



FERRAMENTAS PARA AVALIAÇÃO DA REDE DE MONITORAMENTO DE
QUALIDADE DE ÁGUA DA BACIA DO RIO PIABANHA – RJ COM BASE EM
REDES NEURAIIS E MODELAGEM HIDROLÓGICA

Mariana Dias Villas Boas

Tese de Doutorado apresentada ao Programa de
Pós-graduação em Engenharia Civil, COPPE, da
Universidade Federal do Rio de Janeiro, como
parte dos requisitos necessários para a obtenção do
título de Doutor em Engenharia Civil.

Orientadores: José Paulo Soares de Azevedo
Francisco Olivera

Rio de Janeiro
Março de 2018

FERRAMENTAS PARA AVALIAÇÃO DA REDE DE MONITORAMENTO DE
QUALIDADE DE ÁGUA DA BACIA DO RIO PIABANHA – RJ COM BASE EM
REDES NEURAIS E MODELAGEM HIDROLÓGICA

Mariana Dias Villas Boas

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM
CIÊNCIAS EM ENGENHARIA CIVIL.

Examinada por:

Prof. José Paulo Soares de Azevedo, Ph.D.

Prof. Francisco Olivera, Ph.D.

Prof. Alexandre Gonçalves Evsukoff., Ph.D.

Prof. Celso Bandeira de Melo Ribeiro, D.Sc.

Prof. Cristovão Vicente Scapulatempo Fernandes, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2018

Villas Boas, Mariana Dias

Ferramentas para avaliação da rede de monitoramento de qualidade de água da bacia do rio Piabanha – RJ com base em redes neurais e modelagem hidrológica/Mariana Dias Villas Boas - Rio de Janeiro: UFRJ/COPPE, 2018.

XIV, 153 p.: 29,7 cm.

Orientadores: José Paulo Soares de Azevedo

Francisco Olivera

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Civil, 2018.

Referências Bibliográficas: p. 154-167.

1. Qualidade de Água. 2. Redes de Monitoramento. 3. Redes Neurais. 4. Modelagem hidrológica 5. Bacias Experimentais. I. Azevedo, José Paulo Soares de *et al.* II. Universidade Federal do rio de Janeiro, COPPE, Programa de Engenharia Civil. II. Título.

À minha querida família, em especial,
Aos meus filhos Manuela e Antônio por existirem;
Ao meu marido Leonardo, por seu amor;
Aos meus pais Julio e Alice pelo apoio incondicional;
À minha vó Neyde, pela motivação alegre;
À minha irmã Julia, pela amizade;

AGRADECIMENTOS

O caminho para chegar até aqui foi longo, trabalhoso, desafiador e recheado de boas surpresas. O Doutorado me proporcionou algo maior do que a formação para ser uma pesquisadora, me trouxe mais confiança, mais serenidade e menos cobrança. E chegar até aqui só foi possível graças a pessoas especiais que estiveram presentes fisicamente ou de coração, mas que acima de tudo me fizeram acreditar que era possível.

A minha família, meus filhos, Manuela e Antônio, que já me conheceram nessa loucura de fazer uma "tese" e que me fazem felizes pelo simples fato de existirem. Meu marido, um dos meus maiores motivadores, meu terceiro orientador (essa tese é sua também!), pelo melhor amigo, meu amor. Meus pais, Julio e Alice, pelo apoio incondicional e pelos exemplos de vida que, diariamente, estão presentes nas minhas atitudes e que me fizeram a pessoa que sou hoje. Minha irmã Julia, uma companhia para todo e qualquer momento e uma amiga para toda a vida. A minha avó Neyde, uma incentivadora de marca maior que nunca poupou esforços para me ajudar em tudo, com muito carinho e dedicação. Aos meus amigos e demais familiares que dividiram comigo as angustias, incertezas e as alegrias desse caminho.

Aos meus orientadores, Zé Paulo e Francisco. Zé Paulo, obrigada por acreditar no meu potencial. Seu apoio foi fundamental para eu tentar mediar o racional e a emocional e conseguir fechar essa tese com os "pés mais próximos do chão". Ao professor Francisco que me recebeu na Universidade Texas A & M durante os dois anos do meu doutorado "sanduiche" com quem eu aprendi a confiar mais em mim. Ao professor Alexandre por me apresentar ao mundo da mineração de dados e das redes neurais, umas das muitas surpresas desse caminho, que me encantaram e me fizeram alterar o curso dessa tese para incluí-las. Ao professor Cristóvão e ao Celso pela disponibilidade de participar da minha banca.

Ao Serviço Geológico do Brasil – CPRM e a todos os amigos do DEHID, em especial, a equipe do projeto EIBEX que foram fundamentais para o desenvolvimento desse trabalho. Ao Achilles Monteiro, in memoriam, por todo o suporte e a Lígia Araújo por abrir as portas para o mundo da qualidade de água e das bacias experimentais. A CAPES (MEC/MCTI/CAPES/CNPq) pelo apoio financeiro através do projeto de Pesquisador Visitante Especial.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

FERRAMENTAS PARA AVALIAÇÃO DA REDE DE MONITORAMENTO DE
QUALIDADE DE ÁGUA DA BACIA DO RIO PIABANHA – RJ COM BASE EM
REDES NEURAI E MODELAGEM HIDROLÓGICA

Mariana Dias Villas Boas

Março de 2018

Orientadores: José Paulo Soares de Azevedo.

Francisco Olivera

Programa: Engenharia Civil

O monitoramento da qualidade da água é uma questão complexa que requer ferramentas de suporte para fornecer informações sobre gerenciamento de recursos hídricos. As restrições orçamentais, bem como um projeto de rede inadequado, exigem o desenvolvimento de ferramentas de avaliação para fornecer um monitoramento eficiente. Assim, são propostas e aplicadas duas metodologias para a avaliação da Rede de Monitoramento de Qualidade da Água da Bacia do rio Piabanha (RMQAP). A primeira envolve a análise de componente principal não linear (ACPNL) com base em uma rede neural autoassociativa para avaliar a redundância dos parâmetros e estações de monitoramento da RMQAP. A Análise de Componentes Principais (PCA) é amplamente utilizada para este propósito, entretanto, não captura as não-linearidades características dos dados de qualidade da água, enquanto as redes neurais podem representar essas relações. A partir dos resultados da NLPCA, o parâmetro mais relevante é Coliformes Fecais e o menos relevante é a Demanda Química de Oxigênio. Em relação às estações de monitoramento, a mais relevante é Rocio e a menos relevante é Esperança. A segunda metodologia tem como objetivo a avaliação das estações da RMQAP tendo em vista o impacto dos dados observados na calibração do modelo hidrológico SWAT. Para mensurar esse impacto foi desenvolvido o índice I_{RMQAP} com base no ajuste do modelo hidrológico e de redes neurais para a simulação do parâmetro nitrato em função da vazão em cada estação. Os resultados mostraram que a estação mais impactante é Pedro do Rio e a menos impactante é Poço Tarzan.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

EVALUATION TOOLS FOR THE WATER QUALITY MONITORING NETWORK
OF PIABANHA RIVER WATERSHED BASED ON NEURAL NETWORKS AND
HYDROLOGICAL MODELING

Mariana Dias Villas Boas

March/2018

Advisors: José Paulo Soares de Azevedo

Francisco Olivera

Department: Civil Engineering

Water quality monitoring is a complex issue that requires support tools in order to provide information for water resource management. Budget constraints as well as an inadequate water quality network design call for the development of evaluation tools to provide efficient water quality monitoring. For this purpose, a nonlinear principal component analysis (NLPCA) based on an autoassociative neural network was performed to assess the redundancy of the parameters and monitoring locations of the water quality network in the Piabanha River watershed. Principal Component Analysis (PCA) is widely used for this purpose. However, conventional PCA is not able to capture the nonlinearities of water quality data, while neural networks can represent those nonlinear relationships. From the results of NLPCA, the most relevant water quality parameter is Fecal Coliforms and the least relevant is Chemical Oxygen Demand. Regarding the monitoring locations, the most relevant is Rocio e the least relevant is Esperança. The second methodology aims to evaluate the RMQAP stations in view of observed data impact on the SWAT model calibration. To measure this impact, I_{RMQAP} index was developed based on the adjustment of the hydrological model and neural networks for the simulation of the nitrate parameter as a function of the flow rate. The results showed that the most impressive station is Pedro do Rio and the less impressive is Poço Tarzan.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	CONSIDERAÇÕES INICIAIS	1
1.2	MOTIVAÇÃO E JUSTIFICATIVA	3
1.3	OBJETIVOS	6
1.4	CONTRIBUIÇÃO DA TESE	7
1.5	ORGANIZAÇÃO DO TEXTO	7
2	REFERENCIAL TEÓRICO E BIBLIOGRÁFICO	9
2.1	O MONITORAMENTO DE QUALIDADE DA ÁGUA	9
2.2	O MONITORAMENTO DE QUALIDADE DE ÁGUA NO BRASIL	13
2.3	AS REDES/SISTEMAS DE MONITORAMENTO DE QUALIDADE DE ÁGUA	19
2.4	AVALIAÇÃO DE REDES DE MONITORAMENTO DE QUALIDADE DE ÁGUA	22
2.5	MINERAÇÃO DE DADOS: ANÁLISE MULTIVARIADA E INTELIGÊNCIA ARTIFICIAL	29
2.5.1	REDES NEURAIS ARTIFICIAIS (RNA)	33
2.5.2	ANÁLISE DE COMPONENTE PRINCIPAL	40
2.6	MODELOS HIDROLÓGICOS	48
2.6.1	TIPOS DE MODELOS	50
2.6.2	SELEÇÃO DO MODELO	51
2.6.3	SWAT	54
3	CARACTERIZAÇÃO DA ÁREA DE ESTUDO	60
3.1	CARACTERIZAÇÃO FÍSICA	60
3.2	AS BACIAS EXPERIMENTAIS E REPRESENTATIVAS E O PROJETO EIBEX	64
3.2.1	DEFINIÇÕES	64
3.2.2	CONTEXTO HISTÓRICO	65
3.2.3	O PROJETO EIBEX	68
3.3	CLIMA E REGIME HIDROLÓGICO	73
3.4	CARTOGRAFIA SELECIONADA PARA O ESTUDO	78
3.4.1	MODELO DIGITAL DE ELEVAÇÃO	79
3.4.2	MAPA DE SOLOS	80
3.4.3	MAPA DE USO, OCUPAÇÃO DE SOLOS E COBERTURA VEGETAL	81

4 METODOLOGIA PARA AVALIAÇÃO DE REDES DE MONITORAMENTO DE QUALIDADE DA ÁGUA COM BASE EM REDES NEURAI AUTO-ASSOCIATIVAS

84

4.1	METODOLOGIA	84
4.2	ESTUDO DE CASO	86
4.2.1	SELEÇÃO DOS DADOS	86
4.2.2	PRÉ- PROCESSAMENTO DOS DADOS	89
4.2.3	APLICAÇÃO DA METODOLOGIA	93
4.2.4	RESULTADOS	96
4.2.5	DISCUSSÃO	106

5 AVALIAÇÃO DA RMQAP COM BASE EM MODELAGEM HIDROLÓGIA E REDES NEURAI

5.1	METODOLOGIA	112
5.2	ESTUDO DE CASO	116
5.2.1	MODELAGEM HIDROLÓGICA - SWAT	116
5.2.1.1	SELEÇÃO DO DADOS	116
5.2.1.2	DEFINIÇÃO DE CENÁRIOS DE AVALIAÇÃO DA RMQA	125
5.2.1.3	SIMULAÇÃO E AVALIAÇÃO DOS CENÁRIOS PARA VAZÕES E QUALIDADE DE ÁGUA USANDO O MODELO SWAT A PARTIR DAS MÉTRICAS ESTATÍSTICAS;	126
5.2.2	SIMULAÇÃO DOS DADOS DE VAZÃO PARA PREVISÃO DOS DADOS DE QUALIDADE DE ÁGUA USANDO REDES NEURAI;	141
5.2.3	ESTIMATIVA DO IMPACTO DE CADA ESTAÇÃO A PARTIR DO CALCULO DO <i>IRMQAPe</i>	143
5.2.4	DISCUSSÃO DOS RESULTADOS E AVALIAÇÃO DA RMQAP.	144

6 CONCLUSÕES E RECOMENDAÇÕES

7 REFERÊNCIAS BIBLIOGRÁFICAS

LISTA DE FIGURAS

FIGURA 1 - MAPA COM A DENSIDADE DAS ESTAÇÕES DE MONITORAMENTO DE QUALIDADE DE ÁGUA REPORTADA PELOS PAÍSES MEMBROS, POR RDB, À EEA (FONTE: EEA, 2016).....	12
FIGURA 2 - PONTOS DE MONITORAMENTO DAS REDES ESTADUAIS DE QUALIDADE DA ÁGUA (FONTE: ANA, 2012).	15
FIGURA 3. REGIÕES DEFINIDAS PELO PNQA PARA ESTABELECIMENTO DE PROTOCOLOS DE MONITORAMENTO DE QUALIDADE DA ÁGUA.	18
FIGURA 4. ILUSTRAÇÃO DAS COMPONENTES DE UM SISTEMA DE MONITORAMENTO (ADAPTADO DE WARD <i>ET AL.</i> , 2003).....	19
FIGURA 5 - ETAPAS DE PRÉ-PROCESSAMENTO DE DADOS (ADAPTADO DE HAN <i>ET AL.</i> , 2011).	32
FIGURA 6 – ESTRUTURA BÁSICA NEURAL QUE PROCESSA A INFORMAÇÃO DE ENTRADA EM FUNÇÃO DA DE SAÍDA (ADAPTADO DE ABDI, 1999).....	35
FIGURA 7 – EXEMPLO DE RNA DE “DUAS CAMADAS” OU COMO UMA CAMADA OCULTA (ADAPTADO DE HAN <i>ET AL.</i> , 2011).	36
FIGURA 8 - EXEMPLO DE ARQUITETURA RNAA.	47
FIGURA 9 – EVOLUÇÃO TEMPORAL DOS PRINCIPAIS MODELOS DE QUALIDADE DE ÁGUA E PRINCIPAIS PARÂMETROS MODELADOS (DE PAULA, 2011).	52
FIGURA 10 - LOCALIZAÇÃO DA BACIA DO RIO PIABANHA NO MAPA DE REGIÕES POLÍTICO-ADMINISTRATIVAS DO ESTADO DO RIO DE JANEIRO (FONTE: FUNDAÇÃO CIDE, LABGEO IN FUNDAÇÃO COPPETEC, 2010).	60
FIGURA 11. PERFIL LONGITUDINAL DO RIO PIABANHA (FONTE: ARAÚJO, 2016).	61
FIGURA 12 - DISTRIBUIÇÃO DA POPULAÇÃO POR SETORES CENSITÁRIOS COM CAPTAÇÕES DE ÁGUA E LANÇAMENTOS DE EFLUENTES (ARAÚJO, 2016).	63
FIGURA 13 – CARTA DE SUSCETIBILIDADE A MOVIMENTOS GRAVITACIONAIS DE MASSA E INUNDAÇÃO DO MUNICÍPIO PARA A BACIA REPRESENTATIVA ASSOCIADA ÀS ISOIETAS MÉDIAS ANUAIS (ADAPTADO DE CPRM/2010 E CPRM/2013).	63
FIGURA 14 - PAISAGEM A MONTANTE DA ESTAÇÃO ROCIO, NO MUNICÍPIO DE PETRÓPOLIS, NA BACIA DE MATA PRESERVADA.	69
FIGURA 15 - ÁREA DE CULTIVO NA ENCOSTA E UMA PARTE DO RELEVO BEM ACENTUADO, REGIÃO DO BONFIM, NA BACIA DE USO AGRÍCOLA.	70
FIGURA 16 - ESTAÇÃO LICEU NO CENTRO DE PETRÓPOLIS, NA BACIA DE USO URBANO.	70

FIGURA 17 - LOCALIZAÇÃO DA BACIA DO RIO PIABANHA E DAS BACIAS REPRESENTATIVA E EXPERIMENTAIS. (ADAPTADO DE VILLAS-BOAS <i>ET AL.</i> , 2017)	71
FIGURA 18 – A REDE DE MONITORAMENTO HIDROMETEOROLÓGICA DO PROJETO EIBEX (VILLAS-BOAS <i>ET AL.</i> , 2017).....	73
FIGURA 19 - REGIME DE PRECIPITAÇÕES NA BACIA REPRESENTATIVA DO RIO PIABANHA (FONTE: ARAÚJO, 2016).....	74
FIGURA 20. MAPA DE ISOIETAS DA BACIA HIDROGRÁFICA SOBRE O MODELO DIGITAL DE TERRENO BASEADO NO SRTM 30M (ADAPTADO DE CPRM-2010).	75
FIGURA 21 - TOTAIS ANUAIS DE PRECIPITAÇÃO DAS ESTAÇÕES PARQUE PETRÓPOLIS, LICEU, ROCIO, SÍTIO DAS NASCENTES E VILA AÇÚ (ADAPTADO DE SILVA <i>ET AL.</i> , 2017)	76
FIGURA 22- VAZÕES MÉDIAS MENSAIS DO RIO PIABANHA NA ESTAÇÃO PEDRO DO RIO (58405000) (CPRM, 2017).	77
FIGURA 23. VAZÕES DIÁRIAS COM BASE NOS DADOS OBSERVADOS NA ESTAÇÃO PEDRO DO RIO NO PERÍODO DE 1931-2015 (FONTE: ARAÚJO, 2016).	77
FIGURAS 24 E 25 – SRTM COM RESOLUÇÃO ESPACIAL DE 90M E 30M PARA A BACIA REPRESENTATIVA.	79
FIGURAS 26 E 27 – ZOOM DA REGIÃO DO EXUTÓRIO DA BACIA REPRESENTATIVA PARA DEMONSTRAR A DIFERENÇA DE RESOLUÇÃO ENTRE O SRTM-90M E SRTM-30M RESPECTIVAMENTE.	80
FIGURA 28 – MAPA DE SOLOS PARA A BACIA REPRESENTATIVA (ADAPTADO DE EMBRAPA, 2003).	81
FIGURA 29. MAPA DE USO E COBERTURA DO SOLO DO PROJETO RIO DE JANEIRO PARA A BACIA REPRESENTATIVA NA ESCALA 1:250.000 (ADAPTADO DE JACQUES <i>ET SHINZATO</i> , 2000).....	82
FIGURA 30. MAPA DE USO E COBERTURA PARA A BACIA REPRESENTATIVA, ESCALA 1:25.000. (ADAPTADO DE INEA, 2017)	83
FIGURA 31 - LOCALIZAÇÃO DAS ESTAÇÕES DE QUALIDADE DE ÁGUA SELECIONADAS PARA O ESTUDO E DOS USUÁRIOS DE LANÇAMENTO NA BACIA REPRESENTATIVA (VILLAS-BOAS <i>ET AL.</i> , 2011).....	87
FIGURA 32 – BOX-PLOT PARA OS PARÂMETROS DE QUALIDADE DE ÁGUA.	96
FIGURA 33 - COEFICIENTE DE ASSIMETRIA PARA OS PARÂMETROS DE QUALIDADE DA ÁGUA.	97
FIGURAS 34 E 35 – MATRIZ DE CORRELAÇÃO PARA OS PARÂMETROS DE QUALIDADE DE ÁGUA COM BASE NOS COEFICIENTES <i>R</i> DE PEARSON E <i>RHO</i> DE SPEARMAN RESPECTIVAMENTE.	98
FIGURA 36 – BOX-PLOTS PARA AS ESTAÇÕES DE MONITORAMENTO DE QUALIDADE DE ÁGUA.	102

FIGURA 37 – COEFICIENTE DE ASSIMETRIA PARA AS ESTAÇÕES DE MONITORAMENTO DE QUALIDADE DE ÁGUA.....	102
FIGURAS 38 E 39 - MATRIZ DE CORRELAÇÃO PARA AS ESTAÇÕES DE MONITORAMENTO DE QUALIDADE DE ÁGUA COM BASE NOS COEFICIENTES <i>R</i> DE PEARSON E <i>RHO</i> DE SPEARMAN RESPECTIVAMENTE.....	103
FIGURA 40 - O SCREE PLOT APÓS APLICAÇÃO DO PCA AOS LOCAIS DE MONITORAMENTO.	105
FIGURA 41 – ESTAÇÕES SELECIONADAS PARA OBTENÇÃO DOS DADOS CLIMÁTICOS DE ENTRADA.....	120
FIGURA 42 – VAZÕES DIÁRIAS PARA AS ESTAÇÕES DA BACIA RURAL.	121
FIGURA 43 – VAZÕES DIÁRIAS PARA A ESTAÇÃO DA BACIA PRESERVADA.	122
FIGURA 44 – VAZÕES DIÁRIAS PARA AS ESTAÇÕES DA BACIA URBANA.	122
FIGURA 45 - VAZÕES DIÁRIAS PARA AS ESTAÇÕES PARQUE PETRÓPOLIS E PEDRO RIO.....	122
FIGURAS 46 – CARGAS MENSAS DAS ESTAÇÕES PARQUE PETRÓPOLIS E PEDRO DO RIO PARA O PARÂMETRO NITRATO (NO ₃).	123
FIGURAS 47 - CARGAS MENSAS DAS ESTAÇÕES LICEU, MORIN E ESPERANÇA PARA O PARÂMETRO NITRATO (NO ₃).....	124
FIGURA 48 - CARGAS MENSAS DA ESTAÇÃO ROCIO PARA O PARÂMETRO NITRATO (NO ₃).	124
FIGURA 49 - CARGAS MENSAS DAS ESTAÇÕES POÇO TARZAN, JOÃO CRISTH E POÇO DO CASINHO PARA O PARÂMETRO NITRATO (NO ₃).	124
FIGURA 50 – MAPA DA ÁREA DE ESTUDO COM A DELIMITAÇÃO FINAL EM SUB-BACIAS DO SWAT.....	130
FIGURA 51 – RESULTADO DA SIMULAÇÃO DO SWAT NO SWAT CHECK SEM ALTERAÇÃO DE PARÂMETROS PARA O MÉTODO PENMAN-MONTEITH.....	131
FIGURA 52 - RESULTADO DA SIMULAÇÃO DO SWAT NO SWAT CHECK SEM ALTERAÇÃO DE PARÂMETROS PARA O MÉTODO PRIESLEY-TAYLOR.	132
FIGURA 53 - RESULTADO DA SIMULAÇÃO DO SWAT NO SWAT CHECK SEM ALTERAÇÃO DE PARÂMETROS PARA O MÉTODO HARGREAVES.	133
FIGURA 54 - DIAGRAMA UNIFILAR DA RMQAP E ETAPAS DE CALIBRAÇÃO DO MÉTODO DE CALIBRAÇÃO ÚNICA POR CORES.....	135

LISTA DE TABELAS

TABELA 1 – INFORMAÇÕES DAS ESTAÇÕES DE MONITORAMENTO DO PROJETO EIBEX.	72
TABELA 2 - RELAÇÃO DAS ESTAÇÕES DE QUALIDADE DE ÁGUA DA RMQAP.....	88
TABELA 3 - LISTA DE PARÂMETROS DA REDE DE QUALIDADE DE ÁGUA DA ÁREA DE ESTUDO.....	88
TABELA 4 – EXEMPLO DE MATRIZ DE ENTRADA DE DADOS PARA METODOLOGIA PROPOSTA.	90
TABELA 5- VALORES MÁXIMOS, MÍNIMOS, MÉDIA E MEDIANA PARA OS PARÂMETROS DE QUALIDADE DE ÁGUA SELECIONADOS PARA AS BACIAS EXPERIMENTAIS E A REPRESENTATIVA.	92
TABELA 6 - OS AUTOVALORES E OS PERCENTUAIS DE VARIÂNCIA DEFINIDOS PARA CADA COMPONENTE PRINCIPAL.....	99
TABELA 7 – LOADINGS DAS 13 VARIÁVEIS DE QUALIDADE DE ÁGUA PARA OS QUATRO COMPONENTES PRINCIPAIS APÓS A ROTAÇÃO VARIMAX APLICADA AOS PARÂMETROS DE QUALIDADE DE ÁGUA.....	100
TABELA 8 - PRINCIPAIS ARQUITETURAS DE REDES NEURAIS PARA NLPCA PARA AVALIAÇÃO DE PARÂMETROS DE QUALIDADE DA ÁGUA.....	100
TABELA 9 - O IG (%) CALCULADO PARA OS PARÂMETROS DE QUALIDADE DE ÁGUA.	101
TABELA 10- OS AUTOVALORES E VARIÂNCIAS DO PCA PARA OS LOCAIS DE MONITORAMENTO.	104
TABELA 11 - OS <i>LOADINGS</i> DOS TRÊS COMPONENTES PRINCIPAIS APÓS A ROTAÇÃO VARIMAX PARA OS LOCAIS DE MONITORAMENTO.	105
TABELA 12 - PRINCIPAIS ARQUITETURAS DE REDES NEURAIS UTILIZADAS NA ACPNL/RNAA/IG PARA AVALIAÇÃO DOS LOCAIS DE MONITORAMENTO.	106
TABELA 13 - O IG (%) CALCULADO PARA OS LOCAIS DE MONITORAMENTO.....	106
TABELA 14. OS RESULTADOS DOS MÉTODOS ACP E ACPNL/RNAA/IG PARA OS PARÂMETROS DE QUALIDADE DE ÁGUA.....	108
TABELA 15. OS RESULTADOS DOS MÉTODOS ACP E ACPNL/RNAA/IG PARA OS LOCAIS DE MONITORAMENTO.....	109
TABELA 16 – DEFINIÇÃO DAS CARACTERÍSTICAS DAS CLASSES DO SOLO DA BACIA REPRESENTATIVA.	118
TABELA 17 – CONJUNTO DE PARÂMETROS SELECIONADOS E SUAS RESPECTIVAS FAIXAS DE MEDIÇÃO (ABBASPOUR ET AL., 2007).	127
TABELA 18 – PARÂMETROS MAIS SENSÍVEIS PARA CADA ESTAÇÃO DA RMQAP.....	128
TABELA 19 – PARÂMETROS UTILIZADOS NA ANÁLISE DE SENSIBILIDADE DE QUALIDADE DE ÁGUA (NO ₃ E OD).....	128

TABELA 20 - VALORES DAS MÉTRICAS ESTATÍSTICAS PARA AVALIAÇÃO DO DESEMPENHO DO MODELO.....	136
TABELA 21 - VALORES DE NS PARA A CALIBRAÇÃO DE VAZÕES PARA OS CENÁRIOS DEFINIDOS.....	137
TABELA 22 - VALORES DE PBIAS PARA A CALIBRAÇÃO DE VAZÕES PARA OS CENÁRIOS DEFINIDOS.....	138
TABELA 23 - VALORES DE R^2 PARA A CALIBRAÇÃO DE VAZÕES PARA OS CENÁRIOS DEFINIDOS.....	138
TABELA 24 - VALORES DE RSR PARA A CALIBRAÇÃO DE VAZÕES PARA OS CENÁRIOS DEFINIDOS.....	138
TABELA 25 - VALORES DE NSE PARA A CALIBRAÇÃO DE NO ₃ PARA OS CENÁRIOS DEFINIDOS.....	140
TABELA 26 - VALORES DE PBIAS PARA A CALIBRAÇÃO DE NO ₃ PARA OS CENÁRIOS DEFINIDOS.....	140
TABELA 27 - VALORES DE R^2 PARA A CALIBRAÇÃO DE NO ₃ PARA OS CENÁRIOS DEFINIDOS.....	141
TABELA 28 - VALORES DE RSR PARA A CALIBRAÇÃO DE NO ₃ PARA OS CENÁRIOS DEFINIDOS.....	141
TABELA 29 – AVALIAÇÃO DO DESEMPENHO DAS REDES NEURAIIS COM BASE NAS MÉTRICAS ESTATISTICAS	142
TABELA 30 – ERROS RELATIVOS PADRONIZADOS ENTRE 0 E 1 PARA A CALIBRAÇÃO DE VAZÃO USANDO O MODELO SWAT.....	143
TABELA 31 - ERROS RELATIVOS PADRONIZADOS ENTRE 0 E 1 PARA A CALIBRAÇÃO DE NITRATO USANDO O MODELO SWAT.....	144
TABELA 32 - IMPACTO DAS ESTAÇÕES DA RMQAP ESTIMADO PELO I_{RMQAP}	144

1 Introdução

“As questões de qualidade da água não começaram da noite para o dia e não terminarão da noite para o dia, deve haver uma dedicação no sentido de continuar o monitoramento e a avaliação por muitos anos. O atual foco em resultados pode causar problemas porque os resultados que buscamos em qualidade de água levam tempo e o monitoramento é a única forma de determinar se estamos indo ou não ao encontro de nossos objetivos”. Bob Hirsch, hidrólogo do Serviço Geológico Americano - USGS.

1.1 Considerações Iniciais

A demanda crescente de água, aliada à multiplicidade do seu uso tem provocado crises de escassez, gerando conflitos de várias espécies na bacia hidrográfica. Tais conflitos fizeram com que o planejamento dos recursos hídricos passasse a demandar a consideração, de forma conjunta, de aspectos quantitativos e qualitativos. A escassez de água, já enfrentada por alguns países, não está somente relacionada à quantidade, mas também à qualidade da água. Entretanto, no passado, os recursos hídricos eram apenas pensados sob o ponto de vista da quantidade. Pode-se notar este fato, por exemplo, na disponibilidade de séries de dados históricos: as séries de medições de dados quantitativos apresentam-se em uma quantidade muito superior e com períodos bem mais longos de observação quando comparadas aos dados de qualidade de água.

É possível observar diversos casos de rios que, embora com vazão suficiente para sustentar determinados usos da água, não conseguem fazê-lo em razão da poluição que está intimamente ligada a esses usos e aos usos do solo. Sua principal fonte de poluição, nos rios brasileiros, é o lançamento de efluentes, domésticos e industriais, sem qualquer tratamento ou com tratamento ineficaz (VILLAS-BOAS *et al*, 2008).

A preservação da qualidade da água requer uma gestão efetiva com base na sua caracterização. As redes de monitoramento são as conexões entre a água no meio ambiente e os tomadores de decisão, fornecendo as informações necessárias para essa caracterização (WARD *et al.*, 2003). O monitoramento de qualidade de água consiste na observação e medição padronizada do meio ambiente aquático com determinado objetivo (BEHMEL *et al.*, 2016).

Durante muito tempo houve uma grande preocupação com a coleta de dados e os recursos para o monitoramento eram, em sua maioria, direcionados para essa atividade. Entretanto, nas últimas décadas, os programas de monitoramento vêm sofrendo crescente pressão "para fazer mais com menos" devido à situação econômica mundial (HOROWITZ, 2013). Segundo o autor, os programas bem sucedidos de monitoramento de qualidade de água geralmente possuem um equilíbrio entre a capacidade analítica, coleta de amostras, processamento de dados e recursos disponíveis. Inclusive, com a limitação de recursos, os programas de monitoramento começaram a ser questionados acerca da efetiva informação que estavam produzindo para a sociedade em contraste aos grandes bancos de dados produzidos (WARD *et al.*, 2003).

Além disso, vale observar que o projeto de uma rede de monitoramento de qualidade de água (RMQA) deve ser baseado no objetivo do monitoramento, que deve ser convertido em um protocolo que descreva as variáveis medidas, os pontos e a frequência de amostragem (WARD *et al.*, 2003). No entanto, as práticas usuais não seguem esse princípio e muitas das redes, inclusive no Brasil (ANA, 2012), desenvolveram-se de forma experimental pela inexistência de padrões e procedimentos pré-estabelecidos no passado. Por exemplo, pontos de amostragem foram definidos com base na facilidade de acesso e os parâmetros de qualidade da água foram selecionados de acordo com a capacidade do laboratório (WARD *et al.*, 2003).

Portanto, é necessária a avaliação das RMQA existentes de modo a verificar se elas estão alcançando os objetivos pretendidos de uma forma economicamente viável. Como consequência, desde meados do século passado, várias pesquisas têm sido desenvolvidas para a elaboração de ferramentas de otimização e planejamento de RMQA, de forma a implementar e/ou melhorar a eficiência das redes existentes utilizando uma variedade de técnicas, conforme apresentado na seção 2.4. Dada a relevância da área, em paralelo, nota-se a publicação de uma série de trabalhos que revisam e analisam essas pesquisas, e discutem orientações futuras para a área (WARD, 1996; STROBL *et al.*, 2008; KHALIL *et al.*, 2009; HOROWITZ, 2013; BEHMEL *et al.*, 2016). Entretanto, a maioria dessas técnicas não têm sido implementadas pelos gestores de rede por razões diversas. Algumas são muito gerais, outras muito específicas, ou de difícil implementação, (BEHMEL *et al.*, 2016). E, até os dias de hoje, não há uma prática mundialmente aceita e que se aplique a qualquer

RMQA. Pelo contrário, dado sua conexão com aspectos (por exemplo, padrões de qualidade de água, requerimentos regulatórios, variações de uso e ocupação do solo, etc.) particulares de cada bacia monitorada o mais natural é que técnicas estejam associadas a cada situação (BEHMEL *et al.*, 2016).

Dessa forma, atualmente, é preciso considerar práticas economicamente viáveis que lidem com os altos custos envolvidos na instalação e operação das RMQA, que essas sejam adaptáveis e consigam suportar a grande quantidade de dados, que devem ser convertidos em informação. Nesse sentido, é importante o desenvolvimento de metodologias de avaliação de RMQA que explorem novas tecnologias tais como: sistemas de informações geográficas, sensoriamento remoto, inteligência artificial e, assim, tragam inovação ao tema mesmo que baseadas em técnicas já disponíveis (STROBL *et* ROBILLARD (2008); KHALIL *et* OUARDA, 2009).

1.2 Motivação e Justificativa

A motivação dessa tese envolve três aspectos principais: a problemática do monitoramento de qualidade de água no Brasil, a necessidade do desenvolvimento de uma ferramenta de avaliação para a rede de monitoramento de qualidade de água da bacia do rio Piabanha (RMQAP) e o estudo das técnicas tradicionais e modernas de avaliação da qualidade da água e seu uso na avaliação de redes de monitoramento.

O Centro de Gestão e Estudos Estratégicos – CGEE, vinculado ao Ministério da Ciência e Tecnologia, reuniu uma série de especialistas em qualidade de água e produziu um relatório síntese dentro de uma série chamada “Prospecção Tecnológica em Recursos Hídricos” (CGEE; 2003). Neste documento, é apontada a grande deficiência do setor de qualidade de água no país, principalmente, no que concerne a aquisição, análise e divulgação da informação. Acrescenta, ainda, que “o monitoramento de qualidade da água exige cuidados especiais, pois se espera que os dados representem adequadamente a situação da bacia. As variáveis envolvidas são muitas e a resposta da bacia hidrográfica aos diversos processos que ocorrem na sua superfície possui um grau de aleatoriedade bastante expressivo. O planejamento do monitoramento é importante, pois é muito comum a coleta de uma série de dados e, ao final, extração de poucas informações (...)”.

Existem lacunas a serem preenchidas na decisão de onde, quando e o que monitorar. Em países em desenvolvimento, como o Brasil, é necessário pesquisar qual o processo mais adequado de implantação de redes de monitoramento, uma vez que é imprescindível obter-se a informação desejada a um custo mínimo ou um máximo de informações a um custo pré-estabelecido (SOARES, 2001).

PORTO (2003) aponta que a ampliação das redes de monitoramento brasileiras não deve visar apenas as tradicionais redes operadas hoje por órgãos ambientais ou gestores de recursos hídricos, mas deve também apoiar outras iniciativas para redes menores, às vezes, operadas por projetos. Tendo em vista este objetivo, a Agência de Proteção Ambiental americana (EPA) conseguiu ampliar o monitoramento de qualidade da água para pequenas bacias hidrográficas.

No Brasil, o gerenciamento das RMQA no Brasil é de responsabilidade dos estados, conforme descrito na seção 2.2. As redes de monitoramento estaduais funcionam de forma independente e heterogênea, sem estarem submetidas a uma padronização nacional (ANA, 2012). Nessas redes, geralmente, os dados de qualidade são coletados de forma dissociada dos de quantidade, inviabilizando, por exemplo, a modelagem hidrológica.

A fim de superar essas dificuldades, a Agência Nacional das Águas lançou o “Programa Nacional de Qualidade da Água - PNQA” com o objetivo de promover a cooperação entre os diversos operadores de rede, bem como implementar, expandir e otimizar a distribuição geográfica das RMQA (ANA, 2012). Portanto, o desenvolvimento de ferramentas de avaliação de RMQA, conforme proposto neste estudo, visa contribuir para o PNQA e o avanço da pesquisa na área.

O segundo aspecto motivador da pesquisa refere-se à RMQAP, situada no estado do Rio de Janeiro/BR, que tem como principal objetivo a avaliação do impacto dos usos do solo na qualidade de água conforme, detalhada no Capítulo 3. Embora esteja associada à rede fluviométrica existente, ela foi estabelecida, assim como a maioria das redes no Brasil, sem seguir um protocolo de planejamento adequado. Além disso, a RMQAP (operada feita pelo Serviço Geológico do Brasil – CPRM) sofre constantes restrições orçamentárias, durante as quais há questionamentos sobre a sua eficiência.

Dessa forma, é importante o desenvolvimento de metodologias objetivas de avaliação que servirão de suporte ao seu funcionamento.

Por último, observa-se que, tradicionalmente, a avaliação da qualidade da água, na escala de bacia hidrográfica, é realizada usando duas técnicas: a modelagem e o monitoramento (PARAJULI *et al.* OUYANG, 2013). Os modelos hidrológicos fornecem uma avaliação rápida e econômica das condições da qualidade da água, pois podem simular processos hidrológicos, presentes e futuros, relacionados às condições e usos da bacia (SMITH *et al.*, 1997; LOUCKS *et al.* VAN BEEK; 2017). Já o monitoramento, embora mais dispendioso, permite uma visualização do comportamento do rio através de dados reais. No entanto, a modelagem hidrológica apresenta algumas desvantagens, tais como: a exigência de especificação de um grande número de parâmetros e a necessidade de diversos tipos de dados de entrada ou séries longas de dados observados para a calibração (YAN *et al.*, 2016). Além das incertezas inerentes ao processo de modelagem advindas do próprio modelo, dos parâmetros de calibração e dos dados de entrada (HARMEL, R.D. *et al.* SMITH, 2007). Dada a sua capacidade, a modelagem hidrológica já foi utilizada para avaliação de RMQA conforme será detalhado na seção 2.6.

Além do emprego destas duas técnicas, é possível utilizar sistemas integrados de monitoramento e modelagem baseados em modelos hidrológicos ou, alternativamente, em ferramentas sofisticadas de extração do conhecimento já que, em geral, os dados de qualidade de água são multidimensionais, complexos e não-lineares (Yan *et al.*, 2016). Tais características propiciam o uso de algumas das técnicas de “mineração de dados”, especificamente a análise de componente principal (ACP) e as redes neurais artificiais (RNA). A partir dessas técnicas, pode-se, por exemplo, correlacionar a concentração de poluentes com algum fator mais facilmente mensurável (por exemplo, a vazão). Assim, é possível extrapolar os dados de monitoramento para locais não monitorados ou reduzir o número de estações onde o modelo gera resultados comparáveis à dos medidos (LOUCKS *et al.* VAN BEEK; 2017).

Os métodos de análise multivariada, como a ACP, têm sido amplamente utilizados na avaliação de RMQA, embora apresentem algumas limitações, conforme será discutido no seção 2.5.2. Já as RNA oferecem uma alternativa aos métodos tradicionais e têm atraído atenção considerável dada a sua alta capacidade de

representação de relações desconhecidas lineares e não-lineares (seção 2.5.1). Apesar de as RNA já terem sido utilizadas com sucesso em estudos de qualidade de água, as RNA ainda possuem poucas aplicações para a avaliação de RMQA (STROBL et ROBILLARD, 2008; KHALIL et OUARDA, 2009; YAN *et al.*, 2016)

1.3 Objetivos

Pelo exposto na seção anterior, define-se como objetivo dessa tese o desenvolvimento de metodologias de avaliação de RMQAP, de modo a selecionar as estações e/ou parâmetros com maior relevância.

A primeira metodologia proposta visa determinar a relevância dos parâmetros e estações de monitoramento da RMQAP a partir da análise de componentes principais não lineares (ACPNL) com base em RNA. A segunda metodologia tem como foco selecionar as estações de monitoramento que ofereçam uma maior acurácia do ajuste do modelo hidrológico a partir da definição de um índice impacto baseado em redes neurais. A proposição de metodologias “adaptadas” de técnicas tradicionais visa proporcionar uma maior disseminação das novas metodologias e demonstrar o potencial das redes neurais.

Mais especificamente, a pesquisa aqui proposta tem como principais objetivos:

- Mapear das principais técnicas de avaliação das RMQA, a fim de identificar dentre as mais utilizadas aquelas com maior potencial de aplicação para a área de estudo;
- Rever da problemática das bacias representativas e experimentais, na qual a área de estudo está inserida;
- Caracterizar área de estudo onde serão validadas a metodologia desenvolvida para essa tese, inclusive
- Selecionar estações de monitoramento e parâmetros de qualidade de água da RMQAP com maior relevância;
- Comparar do desempenho da técnica de análise de componente principal não-linear proposta baseada em redes em relação à análise de

componentes principais tradicionalmente utilizada para avaliação de RMQA;

- Avaliar o uso de modelagem hidrológica para avaliação da RMQAP e desenvolver um índice que possa ser utilizado na avaliação das estações da RMQAP;

1.4 Contribuição da tese

A principal contribuição dessa tese está no desenvolvimento e disponibilização de ferramentas para avaliação da RMQAP que, eventualmente, possam suportar o gerenciamento do monitoramento na bacia do rio Piabanha.

A primeira ferramenta desenvolvida tem como foco a avaliação da redundância nos parâmetros monitorados e estações da RMQAP através de uma análise de componentes principais não linear (ACPNL) com base em Redes Neurais AutoAssociativas (AANN) como alternativa à análise de componentes principais (ACP), amplamente utilizada para este propósito.

Tal desenvolvimento visou superar a incapacidade da ACP de detectar o comportamento não linear característico de alguns dos parâmetros da qualidade da água que pode resultar em uma análise deficiente. A ferramenta foi desenvolvida através do método de Influência Geral associado à AANN de forma inédita para quantificar a relevância dos parâmetros e estações.

Além disso, foi desenvolvida uma segunda metodologia de avaliação com base em modelagem hidrológica a partir de metodologias disponíveis. E, então, proposta o uso do modelo de redes neurais, alternativamente ao modelo hidrológico, para seleção das estações de monitoramento mais importantes para a RMQAP tendo como base a acurácia do modelo.

1.5 Organização do Texto

O texto da presente tese foi organizado da seguinte maneira. Inicialmente, são apresentadas as considerações gerais, motivação, justificativas, objetivos e a contribuição da tese, no Capítulo 1. Em seguida, o Capítulo 2 reúne os fundamentos

teóricos e a revisão bibliográfica relativos à pesquisa e às metodologias propostas. É feita uma contextualização da modelagem de qualidade de água, internacionalmente e no Brasil, e são apresentados conceitos e definições relacionados às RMQA. A partir de uma ampla revisão bibliográfica, é possível observar a evolução das técnicas mais utilizadas para a avaliação das RMQA a fim de facilitar o entendimento das técnicas utilizadas nesse trabalho. No Capítulo 3, é feita uma revisão da problemática das bacias representativas e experimentais, na qual a área de estudo está inserida, e a caracterização da área de estudo com a apresentação da base cartográfica utilizada. Após esta revisão, apresenta-se a proposta da primeira metodologia para avaliação de RMQA com base na ACPNL, seguida da sua aplicação ao estudo de caso envolvendo a RMQAP (Capítulo 4). No Capítulo seguinte, é apresentada a segunda metodologia com base nos modelos hidrológicos em contraponto às RNA, bem como a sua aplicação para RMQAP (Capítulo 5). Finalmente, apresentam-se as conclusões e recomendações para trabalhos futuros (Capítulos 6).

2 Referencial Teórico e Bibliográfico

Esse capítulo tem como objetivo realizar uma contextualização do tema monitoramento de qualidade de água e, mais especificamente, da avaliação de RMQA sob o ponto de vista histórico, conceitual e bibliográfico. Além disso, serão apresentadas as técnicas matemáticas selecionadas para esse estudo. Assim, na seção 2.1 são reunidas algumas definições de monitoramento de qualidade de água, seguido por um breve histórico do tema e, então, a situação do monitoramento em alguns países cujos sistemas de monitoramento são referências mundiais. Na seção 2.2, é feito levantamento da situação do monitoramento de qualidade de água no Brasil. Na seção 2.3 são reunidos conceitos sobre as RMQA seguido pela seção 2.4 onde é apresentado o tema avaliação de RMQA e seu estado da arte. A seção 2.5 é o primeiro que trata das técnicas selecionadas para o estudo e reúne informações sobre mineração de dados onde se incluem: as redes neurais e a análise de componentes principais. Por último, na seção 2.6, são abordados os conceitos referentes a modelagem hidrológica.

2.1 O Monitoramento de Qualidade da Água

O monitoramento de qualidade da água pode ser definido como o esforço da sociedade em obter informação quantitativa sobre as características físicas, químicas e biológicas da água através da amostragem estatística (SANDERS *et al*, 1987). A qualidade da água é resultado das atividades da sociedade e do ciclo hidrológico natural que possuem, em parte, natureza estocástica. Consequentemente, pode ser tratada estatisticamente como uma variável randômica (WARD *et al*, 1983). O monitoramento pode ser categorizado de diferentes formas, de acordo com (WARD *et al*, 1983):

- 1) duração - monitoramento de longo ou curto termo;
- 2) tipos de variáveis analisadas – monitoramento químico, físico ou biológico;
- 3) parcela do ciclo hidrológico que se deseja monitorar – monitoramento superficial, subterrâneo, da precipitação, etc.

4) propósito do monitoramento – monitoramento de tendências, *background*, fiscalização, etc.

Segundo CHAPMAN (1996), o monitoramento moderno da qualidade da água teve início na década de 50 com foco em questões gerais sem levar em conta métodos sofisticados. O enfoque do monitoramento estava em coletar dados e não na informação que deveria ser extraída (WARD *et al.*, 2003). Com o passar do tempo, surgiram manuais com metodologias individualizadas, ou seja, para cada tipo de corpo de água (rios, lagos ou águas subterrâneas) ou cada tipo de variável (químicos ou biológicos). Juntamente, foram desenvolvidas abordagens visando sistematizar a operação do monitoramento e recomendando o uso de estatística para analisar os dados e produzir informação com o melhor “custo-benefício” (WARD *et al.*, 2003). Somente nos anos 90 que apareceram guias com orientações mais abrangentes e com uma visão mais detalhada de todo o processo. Entretanto, embora o desafio do monitoramento da qualidade da água venha sendo amplamente abordado na literatura desde a década de 1940, ainda não existe uma estratégia global, holística e prática para suportar todas as fases do monitoramento (BEHMEL *et al.*, 2016).

A definição dos objetivos do monitoramento é fundamental para o bom planejamento de uma rede (WMO, 2008; CHAPMAN, 1996; WARD *et al.*, 2003; HOROWITZ, 2013). Ou seja, as variáveis a serem monitoradas e os métodos a serem utilizados devem ser cuidadosamente selecionados para garantir que os objetivos sejam atendidos tão eficientemente quanto possível. Tradicionalmente, de um modo geral, o principal objetivo da avaliação da qualidade de água era uma simples verificação da sua adequação aos usos pretendidos (CHAPMAN, 1996). Com o aumento da preocupação com a disponibilidade hídrica, os conflitos de uso e o desenvolvimento de técnicas mais sofisticadas, os objetivos se multiplicaram e se particularizaram para cada local e situação.

Nos Estados Unidos, de acordo com WARD *et al.* (2003), inicialmente, os objetivos da avaliação da qualidade de água estavam relacionados à navegação, em manter as rotas navegáveis longe de detritos flutuantes e maus odores. Com o advento da segunda guerra mundial, o foco passou a ser o controle da poluição nas águas superficiais. Não havia guias específicos para estabelecer padrões de planejamento do monitoramento, os gestores aprendiam enquanto executavam. Nos anos 70, o

monitoramento passou a ter como finalidade a compreensão das outras fases do ciclo hidrológico, as pessoas queriam, também, conhecer a qualidade da água subterrânea e avaliar a questão da precipitação ácida, por exemplo. Em 1972, o *Clean Water Act* (CWA) é reorganizado e expandido estabelecendo a estrutura básica para regulação de lançamento de efluentes e dos padrões de qualidade das águas superficiais americanas (EPA, 2016). O CWA atribui a Agência de Proteção Ambiental americana (EPA) a responsabilidade pelo “restabelecimento e manutenção da integridade química, biológica e física das águas nacionais” e, conseqüentemente, pelo monitoramento da qualidade da água. A seguir, houve uma demanda por contabilizar a natureza estocástica da qualidade da água e, assim, incluir um foco estatístico no monitoramento. O que atendia, por outro lado, a questão da falta de informação produzida em comparação a grande quantidade de dados gerados. A estatística poderia ser usada para análise de tendências espaciais e temporais da qualidade da água e então produzir informação útil a população. Na década de 80 e 90, surgem artigos questionando a falta da definição clara de objetivos dos programas de monitoramento. O que é totalmente compreensível visto a forma como o monitoramento foi concebido inicialmente. Atualmente, o monitoramento de qualidade da água nos Estados Unidos é realizado em nível federal e, também, por estados, agências locais, universidades e voluntários sob a regulação da EPA. Os dados são armazenados no *Water Quality Portal* (WQP), portal de Qualidade da Água, através de uma cooperação entre EPA, Serviço Geológico Americano (USGS) e Conselho Nacional de Qualidade de Água (NWQMC).

Uma pesquisa realizada pela Comissão Europeia (CE), em 2012, revelou que a “poluição das águas” é uma das cinco principais questões ambientais que preocupa mais da metade da população na Europa (CE, 2016). Por essa razão a CE definiu como uma de suas prioridades a proteção das águas. E, assim, estabeleceu uma nova estrutura de gestão dos recursos hídricos através da Diretriz 2000/60/EC (WFD) que permeia a sua atuação no âmbito da Política das Águas. Uma das grandes questões da CE era estabelecer de que forma, objetivamente, se daria essa atuação já que a organização engloba diversos países com os seus próprios sistemas de monitoramento anteriores a CE. Em 1995, havia uma grande pressão para uma abordagem mais global da Política das Águas culminando com a organização de uma conferência, em 1996 onde foi proposta a nova estrutura da gestão dos recursos hídricos europeia onde foi definida a

bacia hidrográfica como unidade de gerenciamento denominada distrito (RBD). Assim, definiram-se quatro objetivos para o monitoramento: identificação de áreas com problemas ambientais e de áreas ameaçadas, fornecer informações de que a sociedade está se desenvolvendo de forma sustentável e monitorar ações de recuperação. A Agência Ambiental Européia (EEA) é a responsável por coordenar uma rede de países europeus, que possuem suas próprias redes de monitoramento, para a obtenção dos dados necessários para alcançar os objetivos estabelecidos. Os programas nacionais de monitoramento da qualidade da água dos países membros e o distrital na escala de bacia hidrográfica são as ferramentas mais importantes para o controle da qualidade das águas almejada pela CE através da Diretriz. Esse monitoramento é realizado por diversas autoridades regionais e nacionais. Na Figura 1 é apresentada a densidade das estações de monitoramento de qualidade de água reportada pelos países membros a EEA, por RBD. É possível observar que a densidade varia de 0.5 a 9 estações por cada 1000 km².

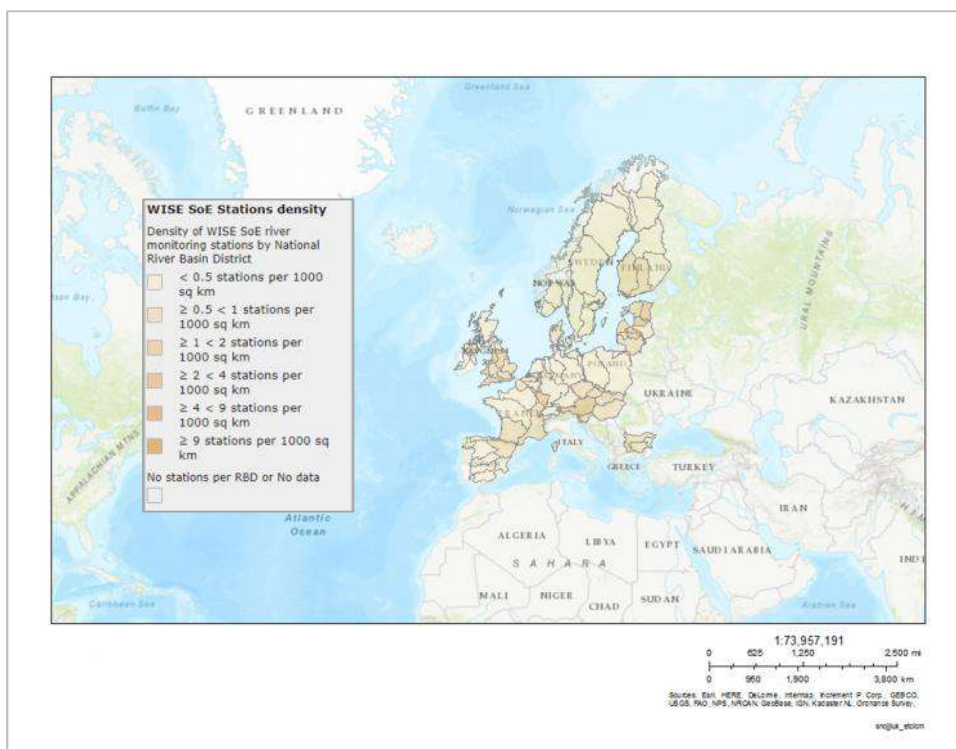


Figura 1 - Mapa com a densidade das estações de monitoramento de qualidade de água reportada pelos países membros, por RDB, à EEA (Fonte: EEA, 2016).

A experiência em monitoramento de qualidade de água da Nova Zelândia é uma referência mundial, principalmente, pois contou com o envolvimento de alguns dos

mais respeitados especialistas da área. A rede de monitoramento nacional iniciou sua operação em 1989 e desde o início é composta por 77 pontos de amostragem, que foram definidos seguindo critérios específicos de projeto como, por exemplo, localizarem-se próximos a estações hidrométricas para ser possível o cálculo da vazão. Ela cobre quase metade do território nacional e é operada por uma única instituição nacional (DAVIES-COLLEY et al., 2011). Destaca-se por ser operacionalmente estável e consistente e, assim, conseguir detectar tendências da qualidade da água em escala nacional (objetivo da Rede) e apoiar campanhas de monitoramento para fins especiais. DAVIES-COLLEY et al. (2011) atribuem essa estabilidade a quatro aspectos principais: a elaboração de um projeto robusto, documentação clara e detalhada, a operação contínua e com poucas alterações na lista de parâmetros, e diversas aplicações dos dados produzidos.

2.2 O Monitoramento de qualidade de Água no Brasil

No Brasil, um país com grandes dimensões e recursos escassos para o monitoramento, as redes de monitoramento da qualidade da água encontram-se instaladas apenas em cerca de dois terços do território, e a maioria delas tem problemas de execução (i.e. períodos longos sem operação). Além disso, observa-se a inexistência de procedimentos de padronização durante o planejamento da maior parte dessas redes, diminuindo, assim, sua eficácia e aumentando os custos operacionais. Dessa forma, as poucas redes de monitoramento, na maioria dos casos, têm funcionado de forma incompleta, deixando de produzir a informação efetiva para ser utilizada, e, ainda, sem um planejamento adequado.

O monitoramento de qualidade de água brasileiro iniciou-se nos anos 70 com a implantação, pelos estados, das primeiras redes de monitoramento de acordo com critérios próprios de planejamento e operação (ANA, 2012). Essas redes estaduais reunidas somavam cerca de 2.167 pontos de monitoramento em 2012 (monitoramento estadual). Além dessas redes, a Rede Nacional Hidrometeorológica passou a incluir a medição de qualidade de água na sua operação e contava, na época, com 1.340 pontos de qualidade de água onde são monitorados cinco parâmetros (temperatura, condutividade elétrica, oxigênio dissolvido, pH e, mais atualmente, turbidez), com uma frequência trimestral ou quadrimestral, de acordo com a operação da rede quantitativa

(monitoramento federal). Durante as medições, que são feitas juntamente com a medição de vazão, é utilizado equipamento portátil multiparamétrico. Há ainda redes de companhias de saneamento (SOARES, 2001).

Considerando os pontos da rede nacional e das redes estaduais, chega-se ao montante de 3.507 estações, que correspondem a uma densidade de monitoramento de aproximadamente 0.4 estação a cada 1000km². Essa atende aquela estabelecida pela Organização Mundial de Meteorologia, que recomenda, em geral, uma densidade de monitoramento que varia de 0,05 a 1 estação a cada 1000 km² dependendo do tipo de região (plana, montanhosa, árida, etc.) (WMO, 2008). Entretanto, se observarmos a Figura 1, com as densidades de monitoramento das bacias da união européia, esse valor ficaria na faixa das bacias menos monitoradas. Vale ressaltar, que essa densidade varia ao longo do território brasileiro, conforme mostra a Figura 2, que apresenta os pontos de monitoramento das redes estaduais. Nota-se que distribuição espacial das estações é bem heterogênea, isto é, enquanto, na região sudeste observa-se uma densidade alta, por exemplo, na região norte a densidade é praticamente nula. A rede de monitoramento paulista, operada pela CETESB (Companhia Ambiental do Estado de São Paulo), apresenta uma densidade que varia de 0,69 a 13,80 pontos a cada 1000 km².



Figura 2 - Pontos de monitoramento das redes estaduais de qualidade da água (Fonte: ANA, 2012).

Vale ressaltar alguns aspectos referentes aos dois principais sistemas de monitoramento que funcionam atualmente: o estadual e o federal. Os sistemas foram concebidos separadamente e funcionam de forma independente, isto é, não há integração e interligação entre as redes federais e estaduais e nem entre as redes estaduais entre si. Dessa forma, é comum que sejam encontradas estações em localizações muito próximas, de entidades diferentes, cada qual monitorando um conjunto de parâmetros em uma frequência muitas vezes coincide (SOARES, 2001). O que significa um desperdício de recursos, pois nesse caso, uma estação poderia ser suficiente. Nota-se, também, uma diferença grande nos níveis de desenvolvimento dos sistemas de monitoramento, há sistemas bem avançados como o do estado de São Paulo, com grande número de estações e protocolo de amostragem bem definidos (SOARES,

2001) e, outros, bem insipientes. Nota-se ainda, na maior parte das redes estaduais, uma dissociação entre monitoramento da qualidade e da quantidade da água fazendo com que raramente haja uma série histórica de vazões associada à série de qualidade. Ao contrário da rede federal, que possui os pontos de monitoramento de qualidade de água localizados junto às estações fluviométricas. Assim, normalmente, se faz a medição de qualidade de água quase simultaneamente a medição de vazão. Por outro lado, a medição de apenas cinco parâmetros feita na rede federal não é suficiente para caracterizar de um modo geral a qualidade de água das bacias. Já na rede estadual observa-se um rol de parâmetros mais extenso.

Com a publicação da Lei 9.433/96, que institui a Política Nacional de Recursos Hídricos (PNRH), essas questões ficaram mais evidentes por irem de encontro a muitos dos princípios da PNRH (BRASIL, 1997). Por exemplo, a definição da bacia hidrográfica como unidade de gestão e implementação da PNRH cujo um dos objetivos seria “assegurar à atual e às futuras gerações a necessária disponibilidade, em padrões de qualidade adequada aos respectivos usos”. Com uma rede nacional com cinco parâmetros não representativos e com redes estaduais com protocolos próprios de monitoramento (por exemplo, frequências de amostragem distintas, elenco de diferentes variáveis, padrões de coleta e análise diversos) que não “conversavam” entre si, ficava inviável definir o padrões de qualidade para determinada bacia hidrográfica. No caso da bacia do Paraíba do Sul, por exemplo, uma bacia federal, seria necessário reunir dados oriundos de das redes de monitoramento do estado do Rio de Janeiro, Minas Gerais e São Paulo com distintos protocolos ou usar os dados da rede federal que não são representativos. Pois, as questões de qualidade de água da bacia são bem complexas, devido à poluição de diferentes origens, tais como: lançamento de esgotos, efluentes industriais, pesticidas, etc. cujos cinco parâmetros monitorados pela rede federal não seriam suficientes. Para definir e avaliar padrões de qualidade seria necessário um rol de parâmetros bem diversificado e específico para essas condições. Outro impasse seria a baixa densidade de estações em algumas bacias, como a Amazônica, muitas vezes insuficiente para a definição de padrões. Além disso, constitui-se como uma das diretrizes gerais para implementação da PNRH a “gestão sistemática dos recursos hídricos, sem dissociação dos aspectos de quantidade e qualidade” o que não ocorre na maior parte das redes estaduais como já relatado.

PORTO (2003) resume bem a situação ao afirmar que um dos maiores déficits na área de qualidade da água no Brasil está na aquisição e utilização da informação, já que além da falta de redes de monitoramento de qualidade da água, a infra-estrutura laboratorial é insuficiente, há dificuldades na análise e divulgação destas informações e existe uma lacuna de capacitação e pesquisa no setor. Entretanto, enfatiza que não basta apenas a definição de novos pontos de monitoramento, o aparelhamento dos estados ou dos laboratórios. É necessário planejamento e o desenvolvimento de procedimentos e diretrizes que irão definir o “protocolo de operação da rede”, Esse protocolo deverá incluir critérios de projeto das redes, diretrizes para uniformização dos procedimentos de coleta e análise das amostras, para análise dos dados e armazenamento das informações.

Tendo em vista os aspectos apresentados e, também, a Lei nº. 10.650/2003, que estabelece que os órgãos ambientais integrantes do Sistema Nacional de Meio Ambiente (SISNAMA) deverão elaborar e divulgar relatórios anuais relativos à qualidade da água, a Agência Nacional de Águas – ANA lançou o Programa Nacional de Avaliação da Qualidade das Águas – PNQA em 2010 (ANA, 2012). O PNQA visa ampliar o conhecimento sobre a qualidade das águas superficiais no Brasil, através de um sistema de monitoramento integrado, com procedimentos de coleta e análise padronizados, que permita um acompanhamento sistemático da evolução da qualidade da água em todo território brasileiro (ANA, 2012). A ideia é definir uma rede nacional com base nas redes estaduais existentes através da cooperação entre os operadores dessas redes, normalmente os órgãos gestores e ANA, denominada Rede Nacional de Monitoramento de Qualidade das Águas (RNMQA). A RNMQA terá como objetivos a análise de tendências, análise da compatibilidade entre a qualidade e os usos da água estabelecidos pelo enquadramento, identificação de áreas críticas, aferição das atividades de recuperação da bacia e apoio ações de planejamento, outorga e fiscalização (ANA, 2014). Para isso dividiu-se o Brasil em cinco regiões de acordo com as características gerais da qualidade da água conforme apresentado na Figura 3. Para cada região foi estabelecido um protocolo de monitoramento que envolve densidade de pontos, frequência de amostragem e parâmetros analisados. Ainda está prevista a ampliação da estrutura de laboratórios, capacitação das equipes técnicas dos estados e divulgação dos dados através do Portal da Qualidade da Água que já está no ar.



Figura 3. Regiões definidas pelo PNQA para estabelecimento de protocolos de monitoramento de qualidade da água.

O PNQA representa um grande passo na direção da melhoria do monitoramento da qualidade da água no Brasil. Mas o desafio é muito grande para sua plena implementação. Já foram desenvolvidas ações de capacitação da equipe e aparelhamento dos estados (ANA, 2014). Apesar da implantação do programa ser feita de forma gradual, ainda há muito a se fazer. A maior parte dos estados brasileiros tem uma estrutura precária para o monitoramento, com equipes com pouquíssimos técnicos, orçamento muito pequeno, instalações inadequadas, falta de veículos para o transporte, dentre outros problemas.

2.3 As Redes/Sistemas de Monitoramento de Qualidade de Água

As redes de monitoramento de qualidade da água (RMQA) podem ser definidas, formalmente, como a localização espacial dos pontos de amostragem (SANDRES et al., 1987). Entretanto, na área de monitoramento de qualidade de água, o termo “redes de monitoramento” é utilizado de forma mais ampla. Em geral, engloba todas as etapas de aquisição do dado de qualidade da água. Assim, quando se fala em projeto ou planejamento de uma rede isto significa a definição da localização dos pontos de monitoramento, dos parâmetros a serem monitorados e a frequência de amostragem com base nos objetivos do monitoramento. WARD et al., vai além e diz que deve-se pensar em “Sistemas de Monitoramento”. O sistema compreende uma série de componentes, ilustrados na Figura 4, alguns relacionados à coleta do dado e outros a geração da informação. Os autores relatam que, historicamente, é dada muita atenção aos três primeiros componentes, referentes à geração dos dados, com isso há muito mais avanço nessa área do que na referente à geração da informação. O tipo de informação gerada dependerá, primordialmente, do objetivo do monitoramento.

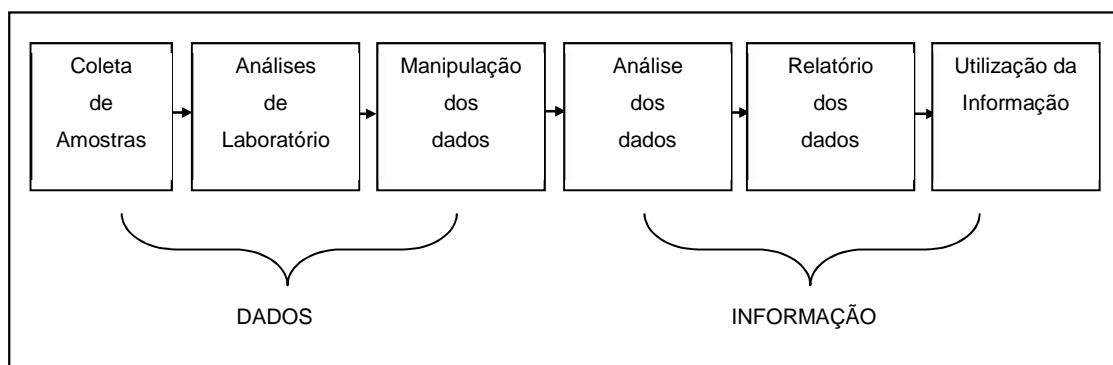


Figura 4. Ilustração das componentes de um sistema de monitoramento (adaptado de WARD *et al.*, 2003)

Dessa forma, o “planejamento de uma rede de qualidade da água” refere-se ao projeto de uma rede numa bacia que não é monitorada ainda. E ele deve ter como base o objetivo do monitoramento de forma a converter esse objetivo em um protocolo que descreva as variáveis medidas, os locais e frequência de amostragem (WARD et al., 2003). Sem essa clara definição, a rede corre o risco de se perder no tempo e no espaço,

e deixar de produzir informação relevante. Já o termo “otimização de uma rede e monitoramento” refere-se ao processo de revisão e aprimoramento de uma rede existente (BEHMEL et al., 2016). Segundo os autores, “otimização” não significa apenas a redução de estações de monitoramento, frequência ou custo. Seu significado é mais amplo e implica na verificação de que a RMQA está atendendo seus objetivos iniciais e se novos objetivos também estão sendo contemplados pelo monitoramento. Os termos avaliação e otimização de RMQA são usados de maneira similar.

STROBL e ROBILLARD (2008) fizeram uma revisão de uma série de projetos de RMQA de forma a investigar os fatores que afetam o desenvolvimento de um efetivo planejamento de uma rede. Eles verificaram que muitas dessas redes foram concebidas casualmente, sem uma estratégia consistente e um planejamento lógico e concluem que o projeto de uma RMQA precisa incluir a combinação dos seguintes fatores: objetivos do monitoramento, pontos de amostragem representativos, as frequências de amostragem, a seleção dos parâmetros qualidade da água e restrições orçamentárias e logísticas.

Na década de 80, teve início o programa internacional de monitoramento e avaliação da qualidade da água “GEMS / WATER” implementado conjuntamente pela Organização Mundial da Saúde (OMS), pela Organização Mundial de Meteorologia (OMM), UNESCO e PNUMA (WHO et al., 1992). O programa ajudou diversos países a estabelecer e fortalecer suas operações de monitoramento da qualidade da água e, ainda, forneceu a eles apoio metodológico. Um dos resultados do programa foi a publicação do “Guia Operacional da Água” que reúne o arcabouço metodológico das práticas empregadas no programa. O Guia traz um passo a passo para o planejamento de uma rede com critérios objetivos para seleção de pontos de amostragem.

Em setembro de 2015, o “Instituto Internacional para o Desenvolvimento Sustentável”, organização canadense sem fins lucrativos, publicou um manual para planejamento de redes de qualidade da água como resultado das questões tratadas no Fórum Econômico Mundial em 2011 (BORDEN et ROY, 2015). O manual consolidou a extensa documentação disponível sobre concepção, metodologia e procedimentos de amostragem, bem como, os demais aspectos que envolvem a concepção de um sistema de monitoramento num documento único. Os leitores são referenciados a documentos com informações mais detalhadas quando necessário. O manual apresenta um quadro

resumo com orientações para o planejamento de uma RMQA que engloba as seguintes etapas:

1. Revisão institucional
2. Identificação dos dados necessários
3. Definição e priorização dos objetivos da Rede
4. Densidade da rede baseada nos objetivos
5. Revisão da rede existente (caso exista)
6. Seleção dos pontos de amostragem e dos equipamentos
7. Selecionar sistema de gerenciamento do dado
8. Estimativa de custo
9. Análise de “custo-efetividade”
10. Implementação

O Manual destaca que a rede precisa ser revista a cada três anos ou num intervalo menor se houver alteração nos objetivos do monitoramento.

No Brasil, ainda vemos poucos esforços na padronização dos procedimentos de planejamento de RMQA. Os órgãos gestores, responsáveis pela operação das RMQA estaduais, de um modo geral, não apresentam documentação sobre seus procedimentos, protocolos e, muito menos, planejamento das suas redes. A documentação clara e detalhada do monitoramento é imprescindível para assegurar estabilidade e consistência da rede (DAVIES-COLLEY et al., 2011).

O PNQA é um grande passo nesse sentido, tanto que a ANA já definiu uma metodologia própria para alocação dos pontos de monitoramento da RNNQA dividida em duas etapas subsequentes: a macrolocação e a microlocação. A macrolocação envolve a identificação das grandes regiões onde deverá ser implementada a rede de monitoramento, e está diretamente relacionada aos objetivos da rede a ser implantada. Já a microlocação envolve a definição precisa dos locais onde o monitoramento deverá ser realizado, onde foi priorizado a localização de lançamento de esgotos doméstico. Existem ainda poucos trabalhos brasileiros nessa linha dos cabe citar, SOARES (2001) que desenvolveu uma metodologia de planejamento e avaliação de redes de

monitoramento baseada no conceito de entropia. Ele faz uma revisão de várias metodologias apresentadas por trabalhos internacionais e afirma que nenhuma delas aponta procedimentos específicos para países em desenvolvimento e COELHO *et* GONÇALVES (2015) que afirmam que é comum a definição de estratégias econômicas de monitoramento e que uma vez definida não são revisitadas.

2.4 Avaliação de Redes de Monitoramento de Qualidade de Água

A avaliação das RMQA é parte fundamental de qualquer sistema de monitoramento como apresentado anteriormente. Mesmo nos casos em que os programas foram estabelecidos de acordo com normas reguladoras e são considerados “adequados”, HOROWITZ (2013) recomenda que seja realizada, frequentemente, uma avaliação da rede de monitoramento, a fim de garantir que estejam funcionando eficientemente e os recursos estejam sendo bem empregados.

De acordo com o Guia Operacional das Águas (WHO *et al.*, 1992), o projeto de uma RMQA é um processo contínuo, com novas estações sendo estabelecidas e as estações existentes sendo descontinuadas à medida que as prioridades e o financiamento evoluem. Assim, deve ser realizado um exame da compatibilidade entre os dados de monitoramento e as informações requeridas após um período adequado a fim de uma utilização mais eficiente dos recursos. Consequentemente, pode ser necessária, por exemplo, a alteração dos locais de amostragem em caso de incompatibilidade.

BORDEN *et* ROY (2015) sugerem uma revisão dos sistemas de monitoramento a cada três anos ou num intervalo menor se houver alteração nos objetivos do monitoramento.

STROBL e ROBILLARD (2008), afirmam que as redes precisam ser periodicamente avaliadas e adaptadas às mudanças das condições ambientais e que novas tecnologias, como a inteligência artificial, devem ser utilizadas para este fim.

De acordo com SANDERS *et al* (2003), qualquer tentativa de avaliar, melhorar ou otimizar uma rede de monitoramento deve começar com a pergunta “Por que nós queremos monitorar?” Isso porque o monitoramento deve ser visto como um sistema

global pelo gestor. Eles destacam que as práticas passadas estiveram voltadas mais para a questão de “como” monitorar ao invés de “por que”.

Dessa forma, nos últimos cinquenta anos, várias pesquisas têm sido desenvolvidas, para a elaboração de ferramentas de avaliação e otimização de RMQA de forma a melhorar a eficiência das redes existentes considerando diferentes características de rede e usando diferentes técnicas matemáticas (por exemplo, Análise de Componentes Principais [ACP], Clustering, Algoritmos Genéticos).

Na década de 70, SHARP (1971) preocupado com o descobrimento de novas fontes de poluição e sabendo que, normalmente, as estações de amostragem de qualidade da água eram posicionadas próximas a municipalidades ou indústrias, de modo a monitorar o lançamento de contaminantes, propõe um plano de amostragem para o Rio Edisto, na Carolina do Sul/EUA, baseado nos centroides dos seus tributários a fim de otimizar a seleção de novas estações. O método é bem incipiente e contém uma parcela grande de subjetividade na definição da macrolocalização.

LETTENMAIER *et al.* (1984) relata a necessidade de otimizar a eficácia dos procedimentos de monitoramento considerando a relação de “custo-benefício” nas decisões de alocação de recursos. E, então propõem uma metodologia para consolidação de redes de monitoramento e expõe a experiência da Agência de Gestão da Qualidade de Água de Seattle, Washington/EUA ao aplicar o método. O método para alocação de estações de monitoramento estabelece uma pontuação ponderada relacionada ao número de ordem dos rios, um pouco similar ao proposto por SHARP (1971). Como resultado, houve a redução do quantitativo de estações da Rede Metropolitana de Seattle, nos Estados Unidos, de 81 para 47 estações. WHITFIELD (1988) discute fatores que influenciam os projetos de RMQA e afirma que os dados precisam ser revisados periodicamente para avaliar a adequação do plano de amostragem.

Na década de 90, a questão custo-benefício continuava em pauta e atrelada a ela surgia a necessidade de avaliação da eficiência das redes de monitoramento existentes face aos novos objetivos que surgiam. Nessa linha, HARMANCIOGLU *et al.* (1992) expõem a dificuldade na avaliação da efetividade e do custo-benefício dos sistemas de monitoramento em face de inexistência de um critério objetivo para ser usado. Para suprir essa necessidade, propõe um procedimento estatístico baseado no

“Princípio da Entropia da Teoria da Informação” que é aplicado a um rio extremamente poluído da Turquia.

Os estudos mais recentes envolvem diversas técnicas tendo em vista, principalmente, a seleção das principais variáveis de uma rede de monitoramento, *i.e.* parâmetros de qualidade da água, pontos de monitoramento e frequência de amostragem. BEHMEL *et al.* (2016) afirma, após uma vasta revisão de cerca de 34 artigos sobre otimização, planejamento e avaliação de redes de monitoramento de qualidade de água, que as diversas críticas feitas aos programas de monitoramento levaram a uma tentativa de padronização dos programas, em todas as suas etapas de modo a fornecer diretrizes e regulamentações para o seu planejamento. Os autores concluem ser extremamente complicado estabelecer uma metodologia única que atenda as diferenças regionais de cada local, tais como: diferença nos requisitos regulamentares, padrões de qualidade da água, diferenças geográficas e geológicas, variações de uso do solo, etc. Por isso, acreditam que até hoje, mesmo com tantas metodologias disponíveis, ainda, não existe uma estratégia prática que atenda todas as fases de planejamento e avaliação de RMQA como já apresentado.

SOARES (2001) propõe uma metodologia para avaliação de desempenho de RMQA com base no conceito de entropia e aplica outros métodos como o de SHARP (1971). O trabalho traz avanços importantes para o desenvolvimento da área no Brasil que possui pouca pesquisa. A metodologia encontra alguns obstáculos na utilização das variáveis de qualidade de água. O conceito de entropia foi utilizado em outras pesquisas conforme relatado por KEUM *et al.* (2017) . PARK *et al.* (2006) combina algoritmos genéticos e sistemas de informações geográficas (SIGs) como uma ferramenta de suporte a decisão útil para otimizar projetos de redes de monitoramento da qualidade da água. A ferramenta define a macrolocalização de pontos de monitoramento usando uma “função-fitness” para o algoritmos genético que requer um pouco mais de pesquisa para ser melhor definida. A pesquisa de CHILUNDO *et al.* (2008) propõe uma RMQA com 16 estações de monitoramento para a Bacia do Rio Limpopo, em Moçambique com base em indicadores químicos, físicos e biológicos. TELCI *et al.* (2009) indicam um modelo de otimização onde os pontos de amostragem são determinados com base na minimização do tempo de detecção de contaminantes utilizando o modelo dinâmico chuva-vazão SWMM. KHALIL *et al.* (2010) desenvolvem uma abordagem para a

seleção otimizada de variáveis de qualidade da água por meio de análise de correlação e *Clustering*. MAHJOURI e KERACHIAN (2011) selecionam a melhor combinação de estações de monitoramento, utilizando algoritmo micro-genético baseado num modelo de otimização e na teoria da entropia para o sistema de monitoramento existente do Rio Jajrood. BEVERIDGE *et al.* (2012) aplicaram duas técnicas de geoestatística (*kriging* e *clustering*) para otimizar redes de monitoramento em grandes lagos. Em conjunto, estas técnicas identificaram estações estatisticamente importantes ou redundantes. CHEN *et al.* (2012) utilizaram modelo numérico de qualidade da água comercial DELFT3D calibrado com os dados disponíveis, a fim de obter projeto ideal e otimizar redes de monitoramento. SCANNAPIECO *et al.* (2012) demonstram o potencial da Análise Fuzzy como ferramenta de otimização na redução da frequência de amostragem. CETINKAYA *et al.* HARMANCIOGLU (2012) aplicam uma aproximação de programação dinâmica modificada como uma ferramenta efetiva para otimizar o número de estações de uma rede de monitoramento. DO *et al.* (2012) introduzem um procedimento de planejamento de redes de monitoramento de forma a identificar pontos de amostragem representativos combinando comprimento de mistura do rio e atividades humanas com o uso de SIGs para localizar os pontos. O trabalho de MAYMANDIA *et al.* (2018) apresenta uma metodologia para otimizar as estações de RMQA de reservatórios e lagos usando o conceito de valor da informação (VOI) e modelagem numérica de qualidade da água e compara os resultados com aqueles obtidos usando uma abordagem teórica de entropia. São utilizados dados de um reservatório do Irã para avaliar a metodologia.

Dentre as diversas técnicas utilizadas para a avaliação das RMQA destacam-se aqui “os modelos hidrológicos” e as “técnicas estatísticas multivariadas”.

Os modelos hidrológicos são muito utilizados na área de qualidade de água para caracterização da situação dos rios. Normalmente as abordagens que usam modelos (sozinhos ou associados a outras técnicas com, por exemplo, análises estatísticas e sistemas de informações geográficas) avaliam as redes de monitoramento sob o ponto de vista da localização das estações. O objetivo é a detecção das áreas mais poluídas da bacia de estudo e que, por isso, requerem um monitoramento mais intensivo. (LO *et al.*, 1996, STROBL *et al.*, 2006, TELCI *et al.*, 2009, ZHU *et al.*, 2018), inclusive, com o modelo SWAT (GIROLAMO *et al.*, 2003). O que não deixa de ser verdade, pois esse é

um dos objetivos mais tradicionais das RMQA, o controle da poluição. Entretanto, é possível utilizar a modelagem como ferramenta de avaliação das RMQA a partir da estimativa do impacto dos dados da estação na calibração do modelo. Pois se existe um modelo que pode representar a bacia razoavelmente, é possível reduzir o número de estações em locais que o modelo apresenta um bom ajuste e que as estações não melhoram a performance do modelo, ou o contrário, aumentar o número onde o ajuste é mais precário.

VANDENBERGHE *et al.* (2002), por exemplo, apresentam uma metodologia para a definição de um conjunto ótimo de dados de amostragem com base na calibração do modelo SWAT (NEITSCH *et al.*, 2011). Um extensivo conjunto de dados de medições produzidos sinteticamente foi reduzido até que a incerteza dos parâmetros do modelo ficasse aceitável. E assim são propostas configurações de rede com menos estações de monitoramento. SMITH *et al.*, (1997) no artigo que descreve o modelo SPARROW (SCHWARZ *et al.*, 2006) afirma que uma área que não deve ser negligenciada para a aplicação do modelo é no planejamento de redes de monitoramento de qualidade de água. O modelo poderia ser utilizado para simulação dos efeitos de mudança na frequência e do local de amostragem com base na melhoria da previsão das saídas do modelo. SAAD *et al.* (2011) avaliam 125.000 estações de monitoramento de qualidade de água ao longo de todo o território americano com base na acurácia dos resultados do modelo SPARROW. Eles observam que os dados de nitrogênio e fósforo de apenas 2.739 estações são indicados para estimativa das cargas dos nutrientes do modelo. A maior parte das estações, que são descartadas pela metodologia, não observa o requisito mínimo do modelo de uma série de 2 anos de dados e 20 amostragens. Além disso, cerca de 50-90% dessas estações não possui dados de vazão associados às concentrações que possam ser utilizados para o cálculo da carga. Assim, concluem que os resultados obtidos, a partir da avaliação dos principais fatores que afetam a acurácia da estimativa das cargas, podem auxiliar a avaliação de redes com a remoção ou exclusão dessas estações, bem como, o planejamento de futuros programas de monitoramento. PURI *et al.* (2017) realizam a otimização da rede de qualidade de água de duas grandes bacias do Texas aplicando algoritmo genético, para uma adequada espacialização do dado e, em seguida, o modelo SPARROW. Os conjuntos ótimos de

estações são aqueles que estimam as cargas de *E. coli* com menos incerteza tendo como referência alguns índices estatísticos.

Por outro lado, o resultado do monitoramento é uma matriz de dados que precisa de uma interpretação complexa para se tornar uma informação útil (SIMEONOV *et al.*, 2003). E as técnicas estatísticas multivariadas são uma ferramenta eficaz na avaliação de bancos de dados grandes e complexos e na conversão de dados em informações (WARD, 2003; SIMEONOV *et al.*, 2003; OGWUELEKA., 2015; HAIR *et al.*, 2009). Além disso, o uso de técnicas estatísticas multivariadas é uma maneira simples e rápida de compreender os processos de qualidade da água quando comparados aos modelos complexos de qualidade hidrológica/água (GIRI *et al.*, 2016). Os modelos de hidrológicos requerem intensivos dados de entrada e conhecimento técnico prévio do comportamento hidrológico da bacia hidrográfica (GIRI *et al.*, 2016).

A Análise de Componente Principal (ACP) é uma das técnicas estatísticas multivariadas mais comumente utilizadas na análise de dados de qualidade da água (KHALIL *et al.*, 2009; OLSEN *et al.*, 2012; GIRI *et al.*, 2016, BEHMELE *et al.*, 2016). Ela tem a capacidade de detectar e eliminar redundâncias, o que, por sua vez, reduz a dimensionalidade do conjunto de dados, mantendo, tanto quanto possível, a variância do conjunto (JOLLIFFE, 2002). Assim, a ACP tem sido frequentemente usada para avaliar a qualidade da água e sua variabilidade espacial e/ou temporal a partir da utilização dos valores de *loadings* e *scores* (VEGA *et al.*, 1998; SIMEONOV *et al.*, 2003; OUYANG *et al.*, 2006; RAZMKHAH *et al.*, 2010; WANG *et al.*, 2012; GUEDES *et al.*, 2012; SELLE *et al.*, 2013). A ACP também permite a avaliação das redes de monitoramento, identificando os parâmetros e locais relevantes que representam a maior variabilidade da qualidade da água (SINGH *et al.*, 2004; OUYANG, 2005; OGWUELEKA, 2015). É indicada para o ajuste da rede, em caso de restrição orçamentária, uma vez que é possível remover os locais e/ou parâmetros de monitoramento redundantes (ou seja, os menos relevantes) sem sacrificar a variância dos dados da qualidade da água (WANG ET *et al.*, 2014; PHUNG ET *et al.*, 2015).

SIMEONOV *et al.*(2003) propõe uma avaliação da origem da poluição em um rio da Grécia a partir da relação entre componentes principais e tipos de poluição. Foram extraídos seis componentes e para cada componente selecionado os parâmetros de qualidade de água mais relevantes (maiores *loadings*). Dessa forma, determinou-se o

tipo de poluição associada a cada componente, quais sejam: por nutrientes, antropogênica, intemperismo, lixiviação, físico-química e orgânica. SHRESTHA *et al.* (2007) fazem uma avaliação da qualidade da água da bacia Fuji, no Japão, de forma similar, identificando parâmetros relevantes a partir dos *loadings*. Nessa mesma linha, PHUNG *et al.* (2015) usam ACP, *clustering* e análise de discriminante para avaliar a estratégia de amostragem para reduzir o número de locais e parâmetros de amostragem na Cidade Can Tho (Vietnã). OUYANG (2005) identifica estações e parâmetros mais importantes da rede de monitoramento do rio St. Johns, na Flórida (EUA), com base nos resultados de análise de fatores e ACP. As estações e parâmetros com maiores *loadings* são as mais importantes para a rede. Ele indica que existe um potencial para melhorar a eficiência da rede reduzindo o número de estações de 22 para 19. WANG *et al.* (2014) sugerem reduzir os custos do programa de monitoramento de água de superfície do rio Tamsui, em Taiwan, eliminando estações de monitoramento redundantes e parâmetros baseados nos *scores* dos fatores. Inclusive, a técnica de ACP já foi aplicada à área de estudo por MOLINARI (2015) de forma a avaliar a RMQA em relação aos parâmetros nitrato e fosfato. A autora concluiu que havia redundância entre as estações, devido ao fato de estarem próximas, e afirmou ser difícil avaliar a bacia como um todo a partir delas.

Os estudos acima mencionados empregaram ACP para avaliar dados de qualidade de água, apesar do fato de que as relações entre os parâmetros de qualidade da água geralmente não são lineares (MCBRIDE, 2005, YAN *et al.* 2016). O resultado é uma avaliação incompleta já que somente as relações lineares são capturadas pelo ACP (WARD *et al.*, 2003; KHALIL *et al.* 2009). Para resolver este problema, recomenda-se a realização de uma análise de componente principal não linear (ACPNL) que pode representar relações tanto lineares como não-lineares entre as variáveis. Na ACPNL, os componentes principais podem ser determinados a partir de uma variedade de métodos, como Redes Neurais AutoAssociativas (RNAA) (um tipo específico de Redes Neurais Artificiais [RNAs]) (KRAMER, 1991), Isomap (TENENBAUM *et al.*, 2000) ou Kernel (SCHOLKOPF *et al.*, 1998), entre outros.

A ACPNL baseada em RNAA (ACPNL/RNAA) foi aplicada com sucesso em diferentes campos, tais como: engenharia, psicologia, compressão de imagens, climatologia, oceanografia, sistemas ambientais, etc. (SINGHAL *et al.* 2000, SALSABURY *et al.* 2000).

2005). Esse método foi selecionado para esse estudo devido à alta capacidade das redes neurais artificiais (RNAs) em ajustar funções não-lineares arbitrárias (KRAMER, 1992, SHARMA *et al.*, 2013). De acordo com MAIER *et al.* (2010), RNAs podem ser facilmente estendidas à problemas de análise multivariada, e parecem ser ideais para modelar as relações entre os parâmetros de qualidade da água dada a sua capacidade de função de aproximação universal vide seu uso em diversos trabalhos de avaliação de dados de qualidade de água inclusive para previsão da parâmetros de qualidade de água em função de dados de vazão (MAIER *et al.* DANDY, 1996; BOWERS *et al.* SHEDROW, 2000; DOGAN *et al.*, 2009; MAIER *et al.*, 2010; NAJAH *et al.*, 2013; SARKAR *et al.* PANDEY, 2015; SARKAR *et al.*, 2015; SEO *et al.*, 2016). Além disso, as redes neurais não requerem um conhecimento prévio acerca das relações entre os dados, diferentemente do método ISOMAP (MAASEN, 2012). Vale acrescentar, ainda, que após vasta revisão bibliográfica foi detectada uma lacuna existente na sua aplicação em estudos de monitoramento da qualidade da água (KHALIL e OUARDA, 2009).

2.5 Mineração de Dados: Análise Multivariada e Inteligência artificial

A mineração de dados é um processo complexo de exploração de grandes bancos de dados a fim de descobrir significativos padrões e regras (BERRY et LINOFF, 2004). Segundo EVSUKOFF (2012), consiste no desenvolvimento de modelos para descoberta de padrões úteis, validos, interpretáveis e previamente conhecidos. Em termos simples, se refere à extração de conhecimento de grandes quantidades de dados (HAN et al., 2011). Também é conhecida pelos termos: extração de conhecimento, análise de padrões, arqueologia de dados, etc. Ela é utilizada em uma ampla gama de aplicações em diferentes áreas, tais como: medicina, biologia, análise de mercado e financeira, gerenciamento de negócios, pesquisa científica, recuperação de imagens, música, etc. (THEODORIS et KOUTROUMBAS, 2009).

A mineração de dados pode ser vista como a evolução natural da tecnologia da informação, resultado, principalmente do aumento do tamanho dos bancos de dados e do desenvolvimento da computação (HAN et al., 2011). Tradicionalmente, a pesquisa por informação em banco de dados era realizada a partir de modelos de descrição que

pressupunham uma anotação manual da informação armazenada (THEODORIS et KOUTROUMBAS, 2009). Tais modelos funcionavam bem a partir de bancos de dados reduzidos e limitados, entretanto, tornaram-se inviáveis diante das grandes massas de dados que passaram a ser armazenadas em sistemas compartilhados por diferentes usuários. Dessa forma, em meados dos anos 80, a área passou por uma grande expansão, impulsionada, ainda, pelo alto poder de computação acessível a uma gama maior de pessoas, inclusive com a disponibilização de softwares comerciais específicos, e a alta demanda por ferramentas que transformassem os dados coletados em informação (BERRY et LINOFF, 2004). Pois, durante muito tempo priorizou-se o desenvolvimento de funcionalidades relacionadas à coleta e ao armazenamento dos dados resultando em grandes quantidades de dados e pouca informação. Essa situação foi qualificada como “Riqueza de dados e pobreza de informações” (HAN et al., 2011). Que pode ser observada em diferentes áreas, como por exemplo, na evolução do monitoramento de qualidade de água apresentada anteriormente, cuja preocupação com a análise e disponibilidade de informação só ocorreu recentemente.

Entretanto, é importante salientar que as técnicas de mineração de dados, em sua maioria, já existiam muito antes da “mineração de dados” se estabelecer com área do conhecimento. Entretanto, elas ficavam restritas a algoritmos acadêmicos ou aplicados a pequenos bancos de dados devido à limitação do poder de processamento da época. Por isso, ela é vista como uma área interdisciplinar que envolve a integração de técnicas de várias disciplinas, como: base de dados e tecnologia de dados, estatística, aprendizagem em máquina, computação de alto desempenho, reconhecimento de padrões, inteligência artificial, visualização de dados, recuperação de informações, processamento de imagem e sinal e análise de dados espaciais ou temporais (HAND, 1998; HAN et al., 2011). BERRY et LINOFF (2004) são enfáticos ao afirmarem que a noção de que mineração de dados e estatística são disciplinas independentes é ultrapassada e que todas as técnicas de mineração de dados tem como base a ciência da probabilidade e estatística. De acordo com FERNANDEZ (2003), todos os métodos de mineração de dados evoluíram a partir de avanços na inteligência artificial, computação estatística e pesquisa de banco de dados e eles não substituem os métodos estatísticos tradicionais, ao contrário, são extensões do uso de técnicas gráficas e estatísticas.

As atividades ou tarefas da mineração de dados dividem-se em dois grandes grupos: predição e descrição. As tarefas descritivas caracterizam as propriedades gerais dos dados e as preditivas fazem inferências dos dados atuais a fim de fazer previsões (HAN et al., 2011). As tarefas de predição podem ser subdivididas ainda em: Classificação e Estimativa; e as de descrição em: Regras de Associação, Clustering e Descrição de perfil (BERRY et LINOFF, 2004). Na Estimativa, a finalidade é prever uma ou mais variáveis de saída em função de uma ou mais variáveis de entrada. Na Classificação os registros são separados em grupos ou classes, previamente definidos, segundo determinadas características. Por outro lado, nos Clustering ou Agrupamento, a tarefa de segmentação que consiste em dividir os registros em grupos mais homogêneos chamados de subgrupos ou clusters anteriormente desconhecidos. No caso das Regras de Associação, a atividade consiste em determinar quais fatos ou objetos tendem a ocorrer juntos numa determinada transação. E na Descrição de Perfil ou Sumarização, o objetivo é descrever de uma maneira simplificada e compacta a base de dados. Ela poder ser realizada numa etapa inicial de qualquer análise de forma a proporcionar um melhor conhecimento da base.

As técnicas de mineração de dados podem ser classificadas, também, de acordo com a forma de aprendizagem em: métodos de aprendizagem supervisionados e não-supervisionados e, ainda, semi-supervisionada (FERNANDEZ, 2003; EVSUKOFF, 2012). Nessa classificação leva-se em conta a utilização ou não de informações de saída pré-definidas na definição dos modelos. Assim para tarefas de predição, normalmente são definidos os métodos de aprendizagem supervisionados ou semi-supervisionados como as Redes Neurais Artificiais (RNA). Já para as de descrição são indicados métodos de aprendizagem não-supervisionados, como por exemplo, a Análise de Componentes Principais (ACP).

O processo de mineração de dados pode ser descrito através dos seguintes passos (HAN et al., 2011):

- 1 – Limpeza dos dados – remoção de ruídos e inconsistências;
- 2 – Integração dos dados – combinação dos dados de diferentes origens;
- 3 – Seleção dos dados – recuperação de dados relevantes
- 4 – Mineração de dados – aplicação da técnica de extração de padrões;

5 – Avaliação de padrões – seleção dos padrões significativos;

6 – Apresentação do conhecimento – são usadas técnicas de visualização e representação do conhecimento para apresentação ao usuário.

As etapas de 1 a 3 integram a fase de chamada “pré-processamento dos dados” e estão ilustradas na Figura 5. Pois, normalmente, as técnicas de mineração de dados são aplicadas a banco de dados reais suscetíveis a erros, valores incomuns, falhas e inconsistências que se não forem cuidadosamente avaliados e corrigidos podem levar a análises ineficientes ou incompletas (HAN et al., 2011). Dessa forma, o conjunto de dados deve ser bem definido e consistente e a sua quantidade deve ser suficiente para suportar a análise (KANTARDZIC, 2011).

A seguir serão apresentadas as técnicas utilizadas nesse trabalho que estão relacionadas à análise multivariada e inteligência artificial. É importante ressaltar que esse trabalho não vai se ater a demonstrações matemáticas das técnicas utilizadas, mas sim, a sua compreensão geral por meio das suas aplicações. Maiores detalhes podem ser encontrados na bibliografia recomendada.

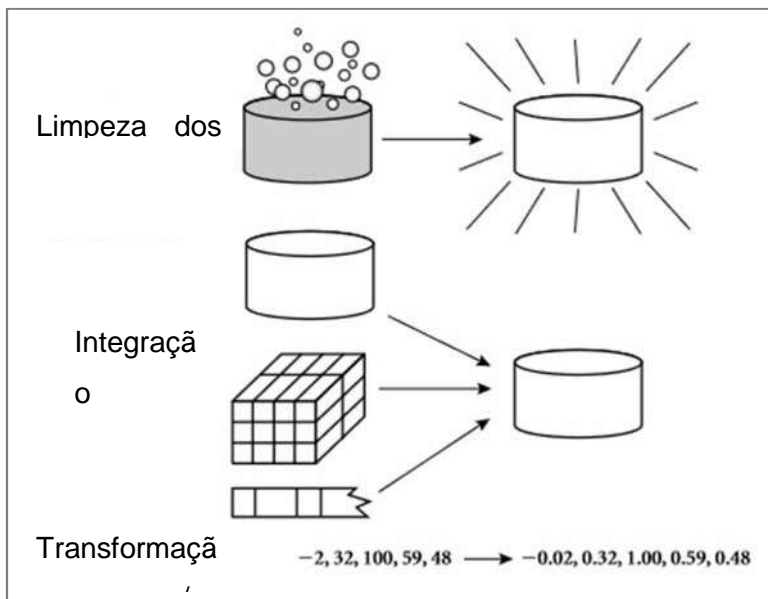


Figura 5 - Etapas de pré-processamento de dados (adaptado de HAN et al., 2011).

2.5.1 Redes Neurais Artificiais (RNA)

A Inteligência Artificial (IA) é uma disciplina relativamente nova. Ela teve início na década de 50, com o advento de computadores acessíveis que transformou as “especulações acadêmicas” das faculdades mentais em uma verdadeira disciplina experimental e teórica (BERRY et LINOFF, 2004). São encontradas diversas definições ao longo do tempo, algumas delas foram reunidas por RUSSELL et NORVIG (1995). Para ilustrar foram selecionadas a mais recente e a mais antiga: "IA é o ramo da informática que se preocupa com a automação do comportamento inteligente" (LUGER et STUBBLEFIELD, 1993 in RUSSELL et NORVIG, 1995) e “IA consiste na automação de atividades que associamos ao pensamento humano, atividades como tomada de decisão, resolução de problemas, aprendizado...”(BELLMAN, 1978 in RUSSELL et NORVIG, 1995). As Redes Neurais Artificiais (RNA) são uma parte fundamental da Inteligência Artificial e sua história, marcada por altos e baixos, se confunde um pouco com a própria história da IA.

As primeiras Redes Neurais artificiais (RNA) surgiram na década de 50 com o advento dos computadores digitais tendo como base o funcionamento dos neurônios biológicos, que já estavam sendo estudados desde 1930 (BERRY et LINOFF, 2004). Por isso são denominadas “artificiais” de forma a contrapor-las às redes do cérebro humano. Houve pouco uso na época em função da limitação do potencial dos computadores e de algumas deficiências teóricas. Somente nos anos 80, com o desenvolvimento do algoritmo Backpropagation (RUMELHART et al., 1986) e sua aplicação a muitos problemas da ciência da computação e psicologia, houve uma disseminação das RNA (BERRY et LINOFF, 2004; RUSSELL et NORVIG, 1995). Mais recentemente, com o seu emprego na área de mineração de dados, o seu uso cresceu e foi expandido para diferentes campos do conhecimento tornando-se uma ferramenta de alto potencial aplicada às áreas da psicologia, estatística, engenharia, economia, marketing etc. (ABDI et al., 1999).

As RNA são modelos estatísticos adaptativos baseados na estrutura do cérebro (ABDI et al., 1999). Como o termo indica, as redes neurais têm uma capacidade de modelagem de inspiração biológica, mas são essencialmente ferramentas de modelagem estatística (RUSSELL et NORVIG, 1995). Elas não diferem, essencialmente, dos modelos estatísticos padrões, pelo contrário, tem a sua teoria embasada em algum deles

(ABDI et al., 1999; FERNANDEZ (2003); BERRY et LINOFF, 2004). Dessa forma, muitos métodos que estão disponíveis em literatura estatística podem e são aplicados a partir das RNA tais como: regressão polinomial, análise discriminante, análise de componente principal, etc. (ABDI et al., 1999; FERNANDEZ, 2003).

O objetivo das redes neurais é “aprender” ou “descobrir”, associações entre os dados de entrada e de saída, através de exemplos de forma similar que o cérebro humano adquire através da experiência (ABDI et al., 1999; FERNANDEZ, 2003; BERRY et LINOFF, 2004). Quanto mais exemplos ou observações (i.e. “mais experiência”) melhor será o ajuste da rede. Por isso é uma técnica de aprendizagem supervisionada como definido anteriormente.

O aprendizado é realizado com base em uma estrutura formado por um conjunto de elementos chamados neurônios interligados e organizados em camadas. Neurônios podem ser definidos como elementos que processam a informação recebida e enviam a informação processada para outros neurônios. A ligação entre os neurônios possui um peso numérico associado. O processamento da informação é feito através de uma função de ativação, definida para cada neurônio, em duas etapas: primeiro é computado a soma ponderada dos dados de entrada e, em seguida, essa soma é transformada através de uma função de ativação. As funções de ativação mais comuns são: a função linear, a função logística e a função tangente hiperbólica (ABDI et al., 1999). A Figura 6 representa a estrutura básica neural onde x_i correspondem aos dados de entrada e w_i aos pesos.

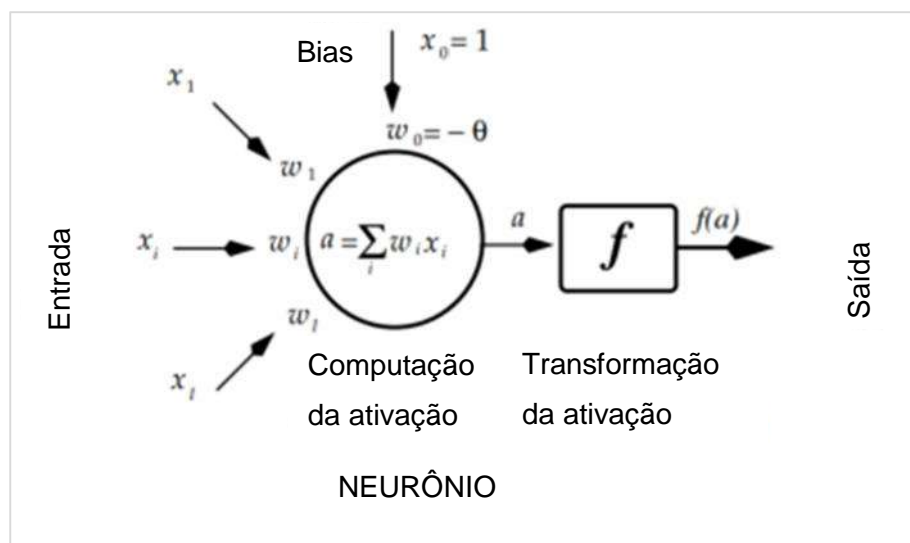


Figura 6 – Estrutura básica neural que processa a informação de entrada em função da de saída (adaptado de ABDI, 1999)

No caso específico de das RNA *feedforward*, comumente utilizadas, cada neurônio na rede recebe informações de neurônios na camada anterior e envia informações para neurônios na próxima camada. A primeira camada é chamada de “camada de entrada” e tem tantos neurônios quanto o número de variáveis de entrada. As camadas seguintes são chamadas de “camadas ocultas” cujo número de neurônios é variável. A última camada é chamada de “camada de saída” e tem tantos neurônios quanto o número de variáveis de saída. O arranjo de neurônios em camadas e sua interconectividade é chamado de “arquitetura ou topologia de rede”. Embora não exista muitas regras para definir a arquitetura de uma RNA, é sabido que essa definição afeta a sua precisão. Entretanto, normalmente, uma rede de feedforward com uma camada oculta pode ajustar qualquer função contínua, e uma rede com duas camadas ocultas pode se ajustar a qualquer função (RUSSELL et NORVIG, 1995). No entanto, o número de unidades em cada camada pode crescer exponencialmente com o número de entradas.

Nota-se que as informações recebidas pela camada de entrada são os valores observados das variáveis de entrada e as informações enviadas pelos neurônios da camada de saída constituem a saída da rede. A camada de entrada não é contabilizada como propriamente uma camada da rede na definição de sua arquitetura, pois não ocorre nenhum tipo de processamento nela, só há uma passagem dos dados de entrada para a próxima camada (BERRY et LINOFF, 2004). Dessa forma, uma rede com uma camada oculta é denominada “rede de duas camadas” como a representada na Figura 7. Nessas redes, a informação enviada por um neurônio é a soma ponderada da informação recebida dos neurônios na camada anterior modificada pela função de transferência mais um termo denominado bias. Os pesos e bias constituem os parâmetros da rede e são ajustados durante o “treinamento”, através de um processo de otimização, que visa a minimização de uma função objetivo. Esse processo é definido através de um algoritmo e corresponde ao “aprendizado da rede”.

O Backpropagation é um exemplo de algoritmo de aprendizagem que funciona a partir de uma RNA feedforward (RUMELHART et al., 1986; HAN et al., 2011). Nele, os pesos e bias são modificados na direção retroativa, ou seja, da camada de saída

através das camadas ocultas até a camada de entrada, a fim de minimizar o erro médio quadrático (EMQ) entre os dados simulados pela rede (saída) e os fornecidos pelo usuário (alvo) dado pela Eq.1:

$$EMQ = \frac{1}{mn} \sum_{p=1}^n \sum_{i=1}^m (a_i - t_i)_p^2 \quad (1)$$

onde m é o número de variáveis; n é o número de observações; a é a saída da rede, e t é o alvo de RNA.

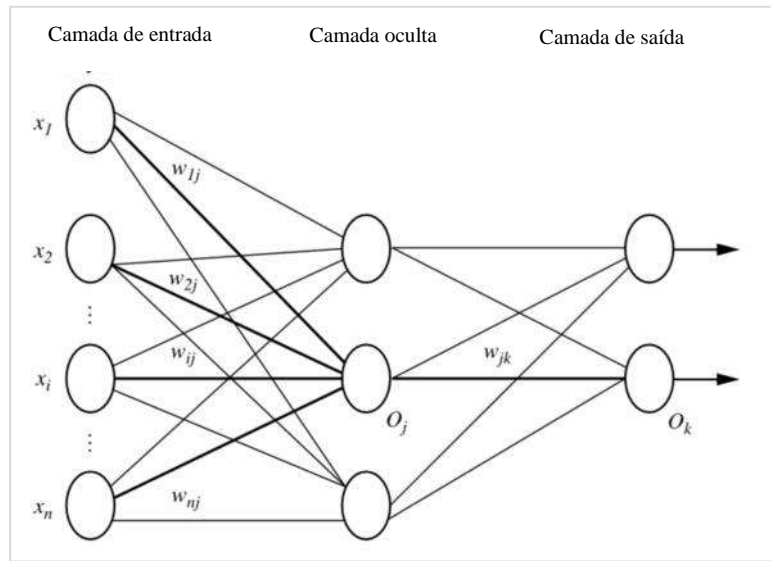


Figura 7 – Exemplo de RNA de “duas camadas” ou como uma camada oculta (adaptado de HAN *et al.*, 2011).

Normalmente, as RNAs são utilizadas na solução de problemas preditivos e descritivos, dada a sua grande capacidade de representar a relação entre as variáveis de entrada e as variáveis de saída. Em termos estatísticos, a interpretação dos parâmetros da rede durante o treinamento pode ser comparada aos valores a e b da equação de regressão linear ($y=a+bx$) (ADBI, 1999). Segundo BERRY et LINOFF (2004) a regressão logística, e mesmo a regressão linear, podem ser vistas como casos especiais de redes neurais.

Assim como em qualquer técnica de mineração de dados, o conjunto de dados utilizado no treinamento de uma RNA deve ser pré-processado. Nota-se, por exemplo, um melhor aprendizado das RNA quando os dados de entrada são mapeados para o intervalo entre -1 e +1. (BERRY et LINOFF, 2004).

O treinamento é um processo de tentativa e erro que deve ser repetido inúmeras vezes em busca do mínimo global. As técnicas de validação e as métricas estatísticas disponíveis podem ser utilizadas para estimar a precisão e validade da rede, como por exemplo, a validação cruzada (RUSSELL et NORVIG, 1995). Isso porque, como todos os modelos estatísticos, as RNA estão sujeitas a overfitting ou falta de generalização. Em outras palavras, uma rede com muitas camadas pode se ajustar perfeitamente aos dados de entrada utilizados para o treinamento, mas poderá falhar em contato com dados novos. Técnicas como a regularização da função objetivo e a parada precoce do treinamento podem ser usadas para evitar esse problema.

Além do overfitting, as RNA possuem algumas desvantagens. A maior seria na representação do seu conhecimento que é de difícil interpretação e, por essa razão, são consideradas “caixas-pretas” (HAN et al., 2011). Algumas pesquisas relacionadas a interpretação dos pesos e bias e análise de sensibilidade estão sendo desenvolvidas motivadas por essa característica da RNA. De acordo com BERRY et LINOFF (2004), elas funcionam melhor quando há apenas algumas variáveis de entrada. Outro inconveniente, é a determinação do número de interações de treinamento. Por isso há alguns métodos disponíveis para isso como, por exemplo, a definição de um valor mínimo para a função objetivo (i.e. quando o usuário tem conhecimento desse valor).

Por outro lado, as RNA apresentam inúmeras vantagens que incluem a sua alta tolerância a dados com ruídos, bem como a sua capacidade de classificar os padrões nos quais não foram treinados. Elas podem ser usadas quando há pouco ou nenhum conhecimento das relações existentes entre os dados. Embora, o aprendizado da rede possa se beneficiar desse conhecimento, caso o usuário o tenha, inclusive na definição da topologia da RNA (RUSSELL et NORVIG, 1995). Elas foram bem-sucedidas em uma ampla gama de dados do mundo real, incluindo o reconhecimento de caracteres, a patologia e a medicina de laboratório (HAN et al., 2011).

BERRY et LINOFF (2004) sugerem os seguinte passo-a-passo para construção de um modelo preditivo usando as RNA:

1. Identificação dos dados de entrada e saída;
2. Pré-processamento com a transformação dos dados no intervalo entre +1 e -1;
3. Configuração da topologia da rede;
4. Treinamento da rede em um conjunto representativo de exemplos de treinamento.

5. Separação dos dados em um conjunto de teste e outro de validação e usar o conjunto de validação para definir os pesos que minimizam o erro;
6. Avaliação da rede usando o conjunto de teste;
7. Aplicação do modelo gerado pela rede para prever resultados para entradas desconhecidas.

Como qualquer modelo a RNA precisam ser atualizadas conforme novos valores de observações forem surgindo. Não são de forma nenhuma uma modelo rígido e imutável (BERRY *et* LINOFF, 2004)

Influência Geral (IG)

Como citado anteriormente, uma das maiores críticas às RNA é o fato de serem vistas como “caixas-pretas”, ou seja, a forma como ocorre seu aprendizado e ajuste entre dados de entrada e de saída são representados implicitamente na forma de pesos e funções. Este conhecimento implícito não está, portanto, diretamente disponível para auxiliar na interpretação e na avaliação da saída da rede, sendo uma limitação ao seu uso como ferramenta de suporte a decisão (HOWES e CROOK, 1999; ALDRICH *et* AURET, 2013). Assim, foram propostas algumas abordagens a fim de entender melhor a dinâmica da RNA. ANDREWS *et al.*, (1995) reúne algumas mecanismos, procedimentos e algoritmos destinados a extrair regras de RNA a fim de fornecer uma visão geral dessas abordagens. BERRY *et* LINOFF (2004) indicam a realização de uma análise de sensibilidade que, embora não tenha regras explícitas, detecta a importância relativa das entradas em relação ao resultado da rede, ou seja, o quão sensível é a saída da rede para cada entrada. Os autores sugerem modificar as entradas a partir de valores médios, mínimos e máximos e avaliar o impacto na saída.

HOWES e CROOK (1999) propõem alguns métodos, dentre eles, um método de estimativa do nível geral de influência de cada variável de entrada na saída de uma RNA com base nos pesos e bias. Eles afirmam que para as RNA não é possível gerar correlações entre variáveis de entrada e saída, como nos modelos de regressão, no entanto, é possível gerar uma estimativa da sua influência denominada “Influência Geral” (IG). O método é semelhante ao desenvolvido por YOON *et al.* (1994),

entretanto, inclui um componente adicional para normalizar para o efeito de pesos extremos e também inclui o termo bias.

Os valores de IG para *feedforward* RNA com apenas uma camada oculta, para cada variável de entrada, com base nos pesos e bias ajustados são calculadas de acordo com a Eq. 2

$$IG(x_i, net) = \sum_{j=1}^{M1} \left(\frac{|w_{ji}|}{\sum_{k=0}^n |w_{jk}|} \sum_{l=1}^f \left| \frac{\hat{w}_{lj}}{\sum_{k=0}^{M1} \hat{w}_{lk}} \right| \right) \quad (2)$$

onde x_i é a i^a variável de entrada (ou seja a $i+2^a$ coluna da matriz de entrada (Tabela 4), *net* refere-se à função de rede neural, w_{ji} é o peso do i^o neurônio de entrada para o neurônio da camada de mapeamento, \hat{w}_{lj} é o peso do j^o neurônio da camada de mapeamento para o l^o neurônio da camada de estrangulamento, M1 é o número de neurônios na camada de mapeamento, e f é o número de neurônios na camada de estrangulamento. Bias são inclusos através do subscrito “0”.

A técnica pode ser utilizada em RNA com apenas uma camada oculta ou múltiplas camadas e tem aplicações em diferentes áreas do conhecimento. PICKERING (2006) investiga a IG de diversos fatores como: democracia, economia interdependência, etc. nas situações de conflitos e não-conflitos. PAPADOKONSTANTAKIS *et al.* (2006) compara quatro métodos de avaliação da influência das variáveis em modelos de dados multivariados usando RNA: Teoria da informação (ITSS), estrutura bayesiana (ARD), influência geral (IG) e omissão sequencial das variáveis (SZW). Os resultados mostram que a IG apresenta o mesmo ranking de variáveis que SZW, e o desempenho dessas duas técnicas, é similar ao da ITSS. Que para o conjunto de dados selecionados, ela é pouco afetada pela arquitetura da rede e seus resultados são melhor distribuídos nos diferentes tamanhos de treinamento em comparação com SZW e ARD, para os quais o tamanho maior dos conjuntos de treinamento parece aumentar o impacto da arquitetura da rede neural. RESINO *et al.* (2011) usa redes neurais para aprimorar o diagnóstico não invasivo de fibrose em pacientes com HIV/HCV. Com base na IG ele estima o peso relativo das variáveis de entrada (*i.e.* características dos pacientes tais como: idade, glicose, colesterol, etc.) na variável de saída (*i.e.* Fibrose significativa e avançada). CHAPMAN

et PURSE (2011) comparam o desempenho de modelos de espécie única e de múltiplas espécies utilizados comumente para avaliar o impacto das mudanças globais nas comunidades ecológicas com base em RNA. Eles calculam a IG de fatores como continentalidade, umidade, solos e urbanização em relação aos modelos para entender a influência de cada um.

2.5.2 Análise de Componente Principal

Nas últimas décadas, as técnicas de análise multivariada têm passado por uma intensa revolução. DEMPSTER (1971) já previa um desenvolvimento notável na área devido ao avanço tecnológico em computação. Além disso, a grande quantidade de informação armazenada em bancos de dados, cada dia mais extensos, que precisa ser interpretada, contribuíram para a ampliação no uso das técnicas de análise multivariada, convertendo o dado em conhecimento. Segundo HAIR JR *et al.* (2009), genericamente, a Análise Multivariada compreende todas as técnicas estatísticas que simultaneamente analisam múltiplas medições, ou seja, qualquer análise simultânea de mais de duas variáveis. Mais precisamente, além de determinar e medir o grau de relação entre variáveis, as técnicas de Análise Multivariada examinam também a relação existente entre a sua combinação. O autor elenca as principais técnicas emergentes e bem estabelecidas de análise multivariada, dentre as quais: a análise de componente principal.

A Análise de Componente Principal (ACP), criada por Karl Pearson em 1901, e posteriormente consolidada por Harold Hottelling em 1933, é uma técnica de análise multivariada que transforma linearmente um conjunto de variáveis, i.e. variáveis originais, em um conjunto de variáveis não correlacionadas, denominadas "componentes" (DUNTEMAN, 1989). Esses componentes são ordenados, de forma decrescente, com base no percentual de variância dos dados representados por cada um deles. Consequentemente, os primeiros componentes explicam a maior parte da variação dos dados originais e são denominados "componentes principais". Neste caso, os componentes são a combinação linear das variáveis originais que representam a maior parte da variação comum do conjunto de dados relativa à correlação linear. Assim, é possível utilizar apenas os componentes principais, ou seja, um número menor de

variáveis, para representar o conjunto de dados ao invés das variáveis originais. Consequentemente, reduz-se a dimensionalidade do conjunto de dados, mantendo-se o máximo possível da variabilidade presente (JOLLIFFE, 2002). Por esta razão, na mineração de dados, a ACP é utilizada muitas vezes na fase de pré-processamento dos dados como um método de redução de dimensionalidade ou remoção de correlações para aplicação de outras técnicas como regressão e *clustering* (HAN *et al.*, 2011). A ACP também pode ser usada para analisar relações inter-relações entre um grande número de variáveis e explicar essas variáveis em termos das suas dimensões subjacentes comuns (HAIR JR *et al.*, 2009).

Os componentes podem ser obtidos através de algumas técnicas dentre as quais a decomposição autovalor-autovetor da matriz de correlação (covariância) dos dados originais, que revela a dimensionalidade do hiperplano através do qual os dados são projetados. A ACP altera as coordenadas do conjunto de dados, projetando as variáveis originais nos sentidos de máxima variância determinada pelos autovalores, que correspondem a matriz de transformação (T) (THEODORIS e KOUTROUMBAS, 2009). Os coeficientes utilizados para gerar as novas coordenadas, denominadas *scores*, são chamados de *loadings*. Eles representam o peso das variáveis originais em cada componente, desde que aquelas estejam na mesma escala, o que pode ser obtido usando algum tipo de procedimento de padronização. Observe que o quadrado de cada *loading* é igual à fração da variância da variável original explicada pelo componente.

De acordo com KRAMMER (1991), ACP é uma fatorização ótima da matriz de dados Y em duas matrizes: T (matriz de dados transformados) e P (matriz de autovetores (*cargas*)) somada a uma matriz de resíduos E . A Eq. 3 representa essa fatorização realizada pela ACP.

$$Y = TP^T + E \quad (3)$$

onde: Y = matriz de conjunto de dados originais; T = matriz de *scores*; P = matriz de *loadings*; E = matriz de resíduos.

Assim, é possível obter os dados originais dos resultados do ACP através da transformação de dados reversa chamada “reconstrução”. Como as linhas da matriz de *loadings* são vetores ortonormais (isto é, a matriz transposta é igual à matriz inversa), os dados originais podem ser obtidos multiplicando a matriz de *loadings* transposta pela

matriz de *scores*. No entanto, como são utilizados apenas os componentes principais a reconstrução de dados não é exata, gerando uma diferença. Assim, é possível avaliar a capacidade de reconstrução de dados do método ACP com base na diferença entre dados originais e reconstruídos. Então, a ACP pode ser vista como o mapeamento ou transformação linear do dado original num novo espaço de coordenadas como representado pela Eq. 4:

$$T = YP \quad (4)$$

A informação perdida nessa transformação pode ser obtida através da reconstrução da matriz de dados representada pela Eq. 5:

$$Y' = TP^T \quad (5)$$

Onde: Y' = matriz de conjunto de dados reconstruída e $Y' = Y - E$

A ACP possui algumas hipóteses e limitações que devem ser investigadas antes da sua aplicação para garantir que os resultados sejam significantes.

- 1) Tamanho das amostras – Sabe-se que quanto maior o conjunto de dados mais efetiva é a análise. Entretanto, há divergências quanto ao tamanho mínimo da amostra que deve ser utilizada (PALLANT, 2010). Idealmente, recomenda-se 150 registros ou pelo menos cinco observações de cada variável.
- 2) Linearidade – ACP assume que as variáveis são linearmente relacionadas. Para isso as correlações entre elas devem ser investigadas a priori e recomenda-se coeficientes de correlação acima de 0.3. Se possível, alguns autores, indicam a realização do Teste de Bartlett's (significância < 0,05) (HAIR JR *et al.*, 2009).
- 3) *Outliers* – ACP é uma técnica sensível a valores espúrios (*outliers*) por isso recomenda-se a sua remoção na fase de pré-processamento dos dados.

A aplicação da ACP consiste, basicamente, das seguintes etapas (HAN *et al.*, 2011) que na maioria dos softwares de mineração de dados é feita de forma automatizada.

- 1) Normalização dos dados de entrada pré-processados de modo que cada variável tenha a mesma escala. Este passo ajuda a garantir que atributos com grandes domínios não dominem atributos com domínios menores e ainda é possível

avaliar a relação entre as variáveis originais e os componentes principais através dos *loadings*.

- 2) Determinação dos componentes através do cálculo dos vetores ortonormais unitários, perpendiculares entre si.
- 3) Os componentes são ordenados por ordem decrescente de "significância" ou força de acordo com o percentual de variação da amostra representado por cada um. Assim, os componentes servem essencialmente como um novo conjunto de eixos para os dados, fornecendo informações importantes sobre variância.
- 4) Como os componentes são classificados de acordo com a ordem decrescente de "significado" o tamanho dos dados pode ser reduzido eliminando os componentes mais fracos, que são aqueles com baixa variação. Teoricamente, usando os componentes principais mais fortes, deveria ser possível reconstruir uma boa aproximação dos dados originais.

Vale notar, que a representação por meio da ACP, como pontuado anteriormente, possui uma grande limitação em face aos dados de qualidade de água: ela pressupõe relações lineares entre as variáveis do conjunto de dados. Assim, quando o objetivo é a análise de dados que podem ter relações não lineares recomenda-se o uso da Análise de Componente Principal não linear (ACPNL).

A ACPNL é uma generalização da ACP e captura tanto relações lineares quanto não-lineares entre as variáveis, onde os componentes são curvas que descrevem a estrutura dos dados em subespaços curvos (KRAMER, 1991). Os primeiros trabalhos que propuseram essa generalização são recentes e datam da década de 80. Podemos reunir quatro principais grupos de pesquisa em ACPNL com base nos métodos empregados: curvas principais e *manifolds*, Kernel, redes neurais e a combinação desses métodos (HSIEH, 2007; KRUGER *et al.*, 2008).

O método das curvas principais (MCP) consistem na flexibilização dos vetores dos *loadings* de forma a encontrar curvas que expliquem a relação entre as variáveis, tratadas simetricamente duas a duas, chamadas de “curvas principais” (HASTIE *et al.* STEUTZLE, 1989). As curvas são obtidas a partir de um algoritmo “intuitivo” e são focadas na distância ortogonal ou menor distância entre os pontos. Já foi demonstrado que MCP e Redes Neurais Auto Associativas são métodos aproximadamente

relacionados (MONAHAN, 2000). *Isomap* (TENENBAUM *et al.*, 2000) é um dos métodos mais utilizados com base em *manifolds* e inclusive já aplicado a dados de qualidade de água (MAASSEN, 2012). O autor analisa os processos dominantes e os fatores de impacto na qualidade de água do wetland Spreewald através do *Isomap*. Mas ressalta que conclusões corretas acerca dos resultados requerem um conhecimento básico da estrutura do sistema e de qualquer processo que possa ocorrer. O algoritmo *Isomap* é um método baseado na abordagem multidimensional clássica de escala onde apenas distâncias Euclidianas são determinadas para calcular o vizinho mais próximo de cada ponto. Ele procura uma incorporação da dimensão inferior mantendo as distâncias geodésicas entre todos os pontos. Essas distâncias geodésicas são organizadas em uma matriz onde será analisada a estrutura não linear dos *manifolds* usando abordagem linear por partes (ROSS *et al.*, 2008).

SCHOLLKOPF (1999) propôs um método com base em funções Kernel (KACP). O método realiza uma análise de componente principal linear e aproxima a função de mapeamento usando as funções Kernel tornando-a não-linear, de forma semelhante a uma rede neural. O que é uma vantagem em relação às redes neurais que encontram, às vezes, dificuldades conceituais e de comunicação para serem usadas (KRUGER *et al.*, 2008). Conceitualmente, consiste em duas etapas; o mapeamento estendido de dados originais no espaço original para o espaço de recursos estendido e o cálculo de ACP no espaço de recursos (CHOI e LEE, 2004). SCHOLZ (2012) comparou RNAA e o KACP e encontrou problemas no reconhecimento de algumas características não-lineares dos dados através do método Kernel.

As redes neurais têm sido extensivamente usadas para a extração de componentes principais não lineares. Os dois métodos mais comuns são as redes neurais baseadas em mapas auto-organizados (SOMs) e as redes neurais auto-associativas (RNAA), utilizadas nesse estudo (ROSS *et al.*, 2008). As redes SOM possuem uma arquitetura específica de forma a projetar os dados de alta dimensão para uma representação discreta de dimensão inferior (geralmente bidimensional), preservando a localidade entre vetores de dados no espaço original de alta dimensão (KOHONEN, 1995). YAN *et al.* (2016) utiliza o ACP e a SOM para analisar um conjunto de dados complexos de estações de monitoramento de água de um rio de Hong Kong. O ACP foi inicialmente aplicado para identificar os principais componentes (PCs) entre os

parâmetros de qualidade da água superficial de comportamento não-linear. Em seguida a SOM foi aplicada para analisar os complexos relacionamentos e comportamentos dos parâmetros através da formação de *clusters*. Entretanto, a aplicação da SOM posteriormente a aplicação da ACP não corrige o inteiramente o problema da detecção não-linearidade da ACP, pois os componentes continuam a ser extraídos apenas com base nas correlação lineares.

O método tradicional de ACPNL baseado em redes Neurais Auto-associativas (ACPNL/RNAA) foi proposto por KRAMER (1991). A ACPNL/RNAA tem sido aplicada com sucesso em diferentes campos e a literatura destaca sua capacidade para encontrar os componentes principais representativos dos dados e para encontrar um mínimo global. MONAHAN (2000 e 2001), por exemplo, detecta um desempenho superior da ACPNL/RNAA, quando comparado com a ACP, para um banco de dados com ruído. Ele testa o dado com várias dimensões e afirma que as diferenças entre as duas técnicas são mais modestas e ocorrem onde os dados são aproximadamente lineares. SINGHAL *et* SALSBUURY (2005) criam um índice de fácil implementação para diagnosticar problemas de controle de válvula devido a presença de correlações não lineares no processo através da ACPNL. Entretanto, ressaltam que esforços futuros devem estar relacionados a melhoria da performance do método. HSIEH (2007), LU *et* PANDOLFO (2011) e SCHOLZ (2012) propõe adaptações ao método tradicional proposto por KRAMER (1991). Os dois primeiros trabalhos focaram em medidas para evitar possíveis problemas nas RNAA, como, *overfitting*. HSIEH (2007) propõe um critério de informação para a definição de um termo de regularização da função objetivo da RNAA e encontrou bons resultados. Já LU *et* PANDOLFO (2011) formulam uma rede neural mais compacta, com apenas duas camadas ocultas e sem o termo bias em duas camadas e obtiveram bons resultados na caracterização dos dados usados. SCHOLZ (2012) propõe um método de aprendizagem hierárquico da RNAA para detectar com mais precisão a não-linearidade dos dados. RAZAVI *et* COULIBAL (2013) comparam os resultados da ACPNL/RNAA (KRAMER, 1991), da sua versão compacta (LU *et* PADOLFO, 2011) e da SOM aplicados a dados de vazão de bacias do Canadá. Os resultados mostraram um desempenho bem similar entre SOM e a versão compacta da NLPCA que não superaram a performance da tradicional. No trabalho de

MIRGOLBABAEI *et al.* (2014), a NLPCA representa adequadamente dados de combustão turbulenta com uma ótima performance na reconstrução desses dados.

KRUGER *et al.* (2008) faz uma análise de alguns desses métodos e afirma que os MCP são conceitualmente simples, mas exigem computacionalmente mais para conjuntos com muitas observações ou muitas variáveis. Com relação aos métodos baseados em redes neurais, os autores afirmam que eles funcionam muito bem, mas são computacionalmente “pesados”, o Kernel, por outro lado, requer uma computação mais leve.

De acordo com KRAMMER (1991) se existe correlação não-linear entre as variáveis de entrada, a ACPNL irá descrever os dados com uma maior acurácia e/ou menos componentes que a ACP. Assim, analogamente a Eq. 4, o mapeamento dos dados na ACPNL pode ser representado pela Eq. 6:

$$T=G(Y) \quad (6)$$

onde: $G=$ é uma função não-linear

A transformação inversa que restaura a dimensionalidade original dos dados é obtida através de uma segunda função não-linear, analogamente a Eq. 5, representada pela Eq. 7:

$$Y'=H(T) \quad (7)$$

onde: $H=$ é uma função não-linear

Assim como na ACP, a perda de informação aqui é medida por: $E=Y-Y'$.

As Redes Neurais Auto Associativas (RNAAs) são um caso particular de RNAs do tipo feedforward treinadas a partir do algoritmo Backpropagation que utiliza os dados de entrada como alvo do modelo (KRAMER, 1991) para realizar um mapeamento de identidade. Ou seja, as RNAAs buscam uma representação dos próprios dados de entrada como uma forma de atender a necessidade de uma formação "supervisionada" requerida pela RNAs.

As RNAAs possuem uma arquitetura específica que permite a execução da ACPNL. Elas contêm cinco camadas, a camada de entrada, a de saída e três camadas ocultas, quais sejam: camada de mapeamento, camada de estrangulamento (ou gargalo) e camada de desmapeamento. A camada de mapeamento é responsável por descrever as

correlações dentre os dados originais, e sua saída é a entrada para a camada de estrangulamento, onde são calculados os componentes principais. A saída da camada de estrangulamento é a entrada para a camada de desmapeamento, que é responsável pela reconstrução dos dados originais. Finalmente, a saída da camada de desmapeamento é comparada com a camada de saída ou alvo (KRAMER, 1991). A Figura 8 exemplifica uma RNAA com cinco variáveis de entrada (I_1 - I_5), três neurônios nas camadas de mapeamento e desmapeamento ($M1=M2=3$) (círculos cinzas) e um neurônio na camada de estrangulamento ($f=1$) (círculo preto). As setas correspondem aos pesos e bias. De acordo com KRAMER (1991), os neurônios das camadas de mapeamento e desmapeamento devem ter funções de ativação não-lineares para que a rede seja capaz de representar relações não-lineares.

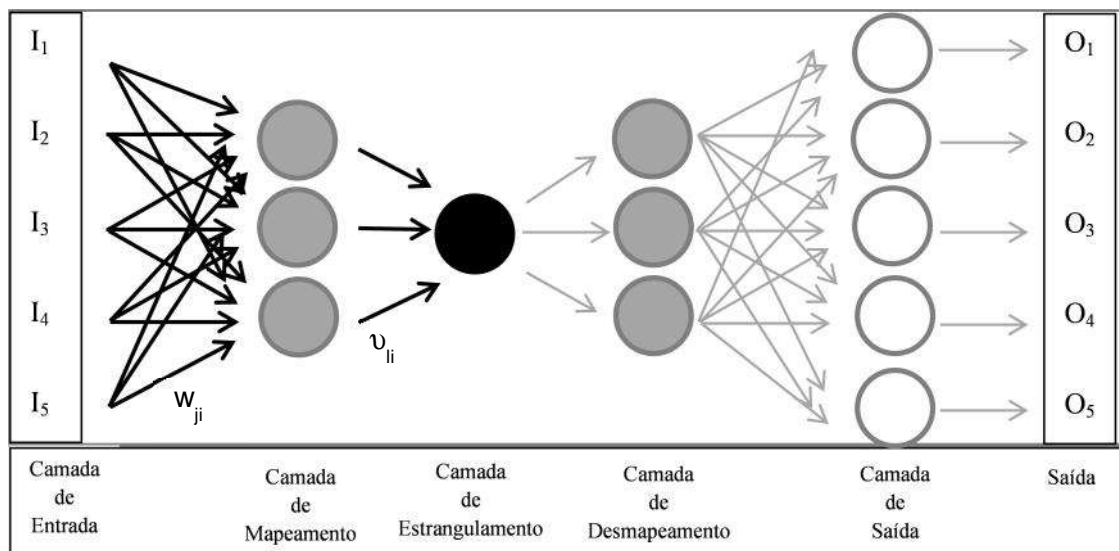


Figura 8 - Exemplo de arquitetura RNAA.

A camada de estrangulamento deve ter necessariamente menos neurônios do que as camadas de entrada e de saída. Essa condição “obriga” a RNAA a desenvolver uma representação compacta dos dados, livre de redundâncias e correlações. Assim, se as informações na camada de entrada puderem ser reproduzidas com sucesso na camada de saída, mesmo passando através da camada de estrangulamento de menor dimensão, significa que os dados de saída da camada de estrangulamento constituem os componentes principais. Consequentemente, o número de neurônios na camada de estrangulamento correspondente ao número de componentes principais.

Dessa forma, a geração das funções G e H (Eq. 6 e 7) através de uma RNAA para a ACPNL é feita da seguinte forma. A função G é representada pelas camadas de entrada, mapeamento e estrangulamento juntas que projetam os dados de entrada num espaço de menor dimensão chamado “espaço fatorial” (KRAMMER, 1991). As camadas de estrangulamento, desmapeamento e de saída representam a função H que a partir dos componentes reproduzem uma aproximação dos dados de entrada. G e H são funções não-lineares por isso as camadas de mapeamento e desmapeamento devem ter funções de ativação não-lineares, permitindo, assim, que a rede seja capaz de captar possíveis relações não-lineares existentes nos dados de entrada (KRAMMER, 1991).

2.6 Modelos Hidrológicos

Os modelos matemáticos são uma tentativa do homem de representar os fenômenos naturais para atender a finalidades específicas. Eles são simplificações da realidade que é traduzida através de parâmetros e equações. O melhor modelo é aquele que agrega pouca complexidade e produz resultados próximos da realidade com o uso de menos parâmetros possíveis (DEVIA *et al*, 2015). Entretanto, a grande questão que envolve os modelos é até que ponto eles podem efetivamente representar a realidade, ou seja, como efetivamente estimar a distância entre a representação e realidade (BEVEN, 2001). Nesse sentido, o estatístico George Box conclui com a célebre frase: “Todos os modelos estão errados, mas alguns deles são úteis” (WASSERSTEIN, 2010).

E, assim, nesse cenário de incertezas de diferentes tipos e graus, o uso de modelos vem crescendo a cada dia apoiado no desenvolvimento da tecnologia computacional que permite a redução do tempo de processamento, a representação de fenômenos variáveis no tempo (por exemplo, modelos dinâmicos) e no espaço (por exemplo, modelos associados a sistemas de informações geográficas), a combinação de dados de diferentes origens, a elaboração de interfaces “hiper” amigáveis, *etc*.

Assim, com a possibilidade de associação de diferentes tipos de dados e tantas outras funcionalidades, os modelos estão cada vez mais sendo usados na área ambiental, mais especificamente na hidrologia. Segundo TUCCI (1998), a representação dos processos hidrológicos através de modelos é a forma encontrada pelo hidrólogo para estudar os diferentes componentes da parte terrestre do ciclo hidrológico relacionados às

interações antrópicas. O autor faz um breve histórico da evolução dos modelos hidrológicos e diz que os primeiros modelos eram fragmentados e tentavam descrever cada componente do ciclo hidrológico, ainda na década 30. Nos anos 70, os modelos envolviam o ajuste de muitos parâmetros e, com o aumento da preocupação ambiental e avaliação do impacto da alteração do uso do solo, iniciou-se o desenvolvimento de modelos com maior base física. A década de 90 é marcada pelo desenvolvimento de modelos climáticos globais e avanços de modelos distribuídos na escala da bacia hidrográfica com avanços importantes aliados ao uso de geoprocessamento. Mais que representar matematicamente o fluxo de água e de seus constituintes, sobre alguma parte da superfície e/ou subsuperfície terrestre, atualmente, há uma preocupação em descrever, também, os processos biológicos e ecológicos e de que forma o regime do fluxo de água pode afetar diversos habitats (ALMEIDA *et* SERRA, 2017).

Os chamados modelos “chuva-vazão” são aqueles que possuem, no mínimo, como variável de entrada a precipitação e como variável de saída a vazão em determinado ponto, ou seja, representam matematicamente a transformação da chuva em vazão. Dependendo da complexidade do modelo, pode haver mais variáveis de entrada (*i.e.* dados de evapotranspiração) e parâmetros (*i.e.* parâmetros que descrevem as propriedades do solo) para descrever esse processo de transformação.

Os modelos hidrológicos são essenciais para a predição e quantificação dos fenômenos físicos que ocorrem na natureza sendo assim ferramentas essenciais para o planejamento e a gestão dos recursos hídricos (ALMEIDA *et* SERRA, 2017). De acordo com TUCCI (2010) as principais aplicações dos modelos hidrológicos são:

- Compreensão do comportamento dos processos hidrológicos;
- Análise de consistência e extensão de séries hidrológicas em locais com poucas informações;
- Dimensionamento e planejamento do desenvolvimento numa bacia hidrográfica;
- Previsão de vazões;
- Simulação de cenários futuros com base em modificações naturais e antrópicas da bacia hidrográfica.

2.6.1 Tipos de modelos

Os modelos “chuva-vazão” podem ser classificados com base em diversos aspectos, tais como (DEVIA et al, 2015; ALMEIDA e SERRA, 2017; TUCCI, 1998) :

- Variação dos parâmetros do modelo no espaço – modelos concentrados ou distribuídos;
- Variação dos parâmetros do modelo no tempo e no espaço – modelos estáticos ou dinâmicos;
- Tipos de variáveis representadas – modelos estocásticos ou determinísticos;
- Tipo de relações entre as variáveis – modelos empíricos, conceituais ou físicos;
- Quanto à escala de representação – detalhados ou genéricos.

Nos modelos concentrados ou agrupados toda a área da bacia é tomada como uma única unidade, ou seja, as saídas são geradas sem consideração da variabilidade espacial. Por outro lado, nos distribuídos, a bacia é dividida em unidades, onde serão estabelecidos os parâmetros do modelo, conectadas através de relações que serão modeladas.

Com relação à variação temporal, têm-se os modelos dinâmicos, nos quais as equações do modelo são resolvidas a cada passo de tempo, e os estáticos onde não há variação temporal dos dados de entrada e saída, pois o modelo processa os dados na forma de passo de tempo único.

O modelo determinístico é aquele que gera a mesma saída para o mesmo conjunto de dados de entrada, enquanto que no caso dos modelos estocásticos, diferentes valores de saída podem ser produzidos para um único conjunto de entradas.

Outra classificação está relacionada às variáveis do modelo. Modelos empíricos são aqueles que utilizam relações para a simulação dos fenômenos baseadas em observações. Por outro lado, quando os processos que envolvem o fenômeno a ser representado são descritos o modelo ele é denominado conceitual. Já os modelos físicos

possuem uma representação matematicamente idealizada do fenômeno real, com os princípios físicos do processo expressamente descritos.

Por fim, o modelo pode ser detalhado de modo a ser apropriado a pequenos intervalos de tempo ou áreas, ou genérico, podendo representar apenas grandes área ou longos intervalos de tempo.

2.6.2 Seleção do Modelo

A definição do modelo apropriado à determinada situação envolve, principalmente, o conhecimento do modelador em relação à compatibilidade entre as características da área de estudo e o tipo de modelo. Segundo TUCCI (2010), a capacidade do modelo para descrever os processos envolvidos depende das formulações utilizadas e de suas limitações. Destaca que é importante conhecer as limitações de forma a diferenciar a fonte dos erros originadas do modelo das demais. Por exemplo, erros gerados pela deficiência dos dados de entrada ou pela inadequada estimativa dos parâmetros.

BEVEN (2001) discute de forma aprofundada as limitações dos modelos hidrológicos agrupando-as em cinco principais aspectos. O primeiro seria a não-linearidade de diversas relações existentes na hidrologia que são tratadas por muito modelos como lineares. A segunda, seria referente à escala do modelo em relação a área de estudo. Acrescenta que os modelos são, em sua maioria, generalistas e não capturam as singularidades de cada bacia. O quarto aspecto abordado relaciona-se à da equifinalidade. Nesse caso, discute-se como selecionar o conjunto de valores dos parâmetros do modelo dentre os vários conjuntos que resultam em um mesmo ajuste na calibração do modelo. Por último, o autor discute as diversas fontes de incerteza inerentes a modelagem.

Dessa forma, é possível perceber que embora, atualmente, muitos modelos disponham de ferramentas que facilitam a sua manipulação e aplicação, elas não excluem o profundo conhecimento do modelador com relação ao mesmo, pelo contrário, podem ser uma armadilha para um usuário leigo. De tal forma que se as limitações do modelo não forem consideradas, avaliadas e destrinchadas o processo de

modelagem hidrológica tende a gerar resultados não confiáveis e imprecisos. Nesse sentido, TUCCI (2001) afirma que a engenharia tem utilizado com parcimônia muitos dos modelos para gerenciamento dos recursos hídricos, mas muitas vezes, sem um exame adequado das suas limitações, o que tem produzido incertezas nas decisões e nos projetos de recursos hídricos. PARAJULI *et* OUYANG (2013) fazem uma revisão de alguns dos principais modelos de qualidade de água e afirmam que a modelagem de hidrologia, de sedimentos e nutrientes desenvolveu-se substancialmente, mas os avanços nem sempre foram consistentes com a compreensão dos usuários com relação aos potenciais, às limitações e às aplicações adequadas de cada um.

A Figura 9 apresenta um quadro com alguns dos principais modelos de qualidade de qualidade de água ao longo do tempo. A seguir há a descrição de alguns modelos hidrológicos livres com potencial de uso nesse trabalho cujo objetivo geral é a avaliação da RMQAP. Sabendo-se que a avaliação de uma RMQA deve considerar, primordialmente, o objetivo da rede de monitoramento e dado que o objetivo da RMQAP é a análise do impacto dos diferentes usos e ocupação da bacia na conjuntura das bacias experimentais e representativas (conforme será detalhada no Capítulo 3). O modelo hidrológico selecionado deveria ser um modelo físico, distribuído, dinâmico que envolvesse a simulação de parâmetros de qualidade de água.

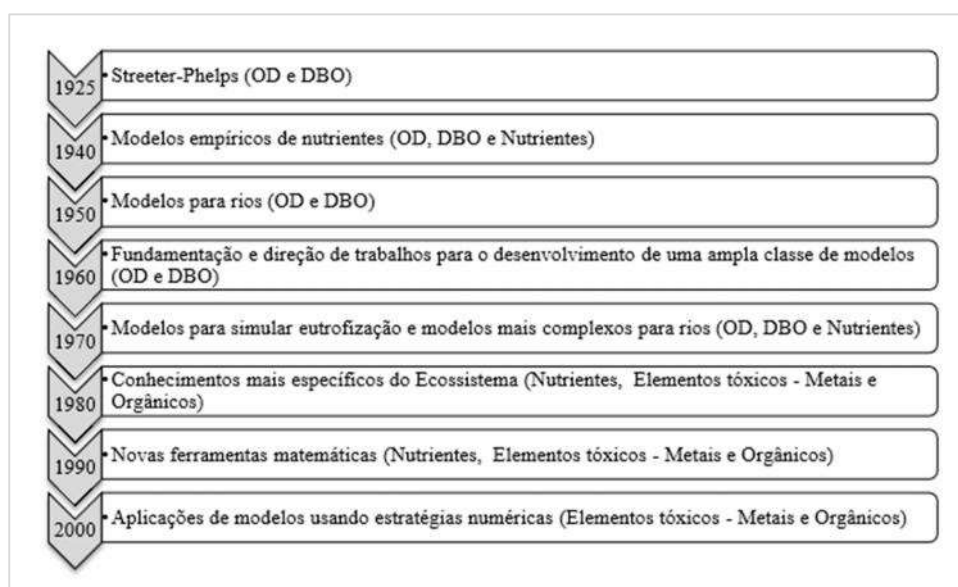


Figura 9 – Evolução temporal dos principais modelos de qualidade de água e principais parâmetros modelados (DE PAULA, 2011).

- SPARROW

O Spatially Referenced Regression On Watershed attributes - SPARROW (SMITH *et al*, 1997) é um modelo de regressão espacial para caracterização da qualidade de água de bacias hidrográficas desenvolvido pelo Serviço Geológico Americano. A modelagem é feita com base nos dados de estações de monitoramento de vazão e concentração de parâmetros de qualidade de água. O modelo possui uma equação de regressão não-linear que descreve o transporte não-conservativo de contaminantes pontuais e de fontes difusas em superfícies terrestres para riachos e rios (SMITH *et al*, 1997). SPARROW é um modelo estatístico livre de escala anual que envolve a simulação de diversos parâmetros de qualidade de água (por exemplo, nutrientes, pesticidas, *e. coli*.). Ele tem sido usado para avaliar hipóteses alternativas sobre fontes importantes de contaminantes e propriedades de bacia hidrográfica que controlam a carga de contaminantes e o transporte em grandes escalas espaciais. Foram realizadas algumas aplicações de SPARROW para avaliar redes de monitoramento de qualidade da água nas bacias hidrográficas nos últimos anos (SMITH *et al*, 1997; SAAD *et al*, 2011; ALAM *et* GOODALL, 2012; PURI *et al*, 2017) descritas na seção 2.4.

- HSPF

HSPF é um modelo hidrológico da Agência Ambiental americana (EPA) de ampla escala, conceitual e dinâmico, que simula vazões de fontes não-pontuais e a qualidade da água associadas a contribuições de fontes pontuais numa bacia hidrográfica (BICKNELL *et al*. 2001). O modelo HSPF usa dados de entrada, tais como: dados de precipitação horária, temperatura e radiação solar, condições de uso do solo; e práticas de gestão de terras para prever parâmetros nas escalas das bacias hidrográficas (PARAJULI *et* OUYANG, 2013). Os resultados das simulações incluem escoamento superficial, carga de sedimentos, nutrientes e concentrações de pesticidas. Não foram encontradas aplicações para avaliação de redes de monitoramento de qualidade da água.

- SIAQUA-IPH

No Brasil, o principal candidato para esse estudo seria o modelo MGB-IPH que possui acoplado o Programa Simulador Analítico de Qualidade da Água (SIAQUA-

IPH) que é um modelo de qualidade de água desenvolvido para simular diferentes cenários de impacto de lançamentos de efluentes em rios com grandes bacias hidrográficas (> 3.000 km²) em situações de carência de dados (MAINARDI, F., 2013).

O MGB-IPH é desenvolvido pelo Instituto de Pesquisas Hidráulicas (IPH) da Universidade Federal do Rio Grande do Sul (UFRGS) e utiliza dados de precipitação, temperatura do ar, umidade relativa, velocidade do vento, insolação e pressão atmosférica para calcular as vazões dos rios de uma bacia hidrográfica na qual a bacia hidrográfica é dividida em pequenas sub-bacias, denominadas mini-bacias, utilizando as técnicas do conjunto de ferramentas do ArcHydro (COLLISCHONN *et al.*, 2010).

Para esse estudo, dentre os modelos hidrológicos disponíveis foi selecionado o modelo SWAT (SWAT, 2012). O modelo tem uma interface amigável, possui diversas aplicações bem sucedidas em pequenas bacias, inclusive no Brasil (ZHANG *et al.*, 2008; QIU *et al.*, 2012; BRESSIANI, 2016) (como a bacia representativa da bacia do Rio Piabanha), possui um código livre e aberto e já foi utilizado em parte da área da bacia com bons resultados (SILVA, 2014). Além disso, o SPARROW, por exemplo, requer cargas anuais de pelo menos 2 anos e 20 amostragens e área de estudo ainda não tem esse quantitativo disponível. Já HSPF exige dados de precipitação horária, ainda não disponíveis para a área de estudo, e exige grandes quantidades de dados de entrada, além de ter uma interface menos amigável (IM *et al.*, 2003). Por outro lado, tendo em vista a escala de interesse do modelo, o SIAQUA-IPH é um modelo para grandes bacias que pode ser aplicado a áreas menores. Entretanto, o modelo é recente e não foram encontradas aplicações em pequenas bacias. Vale acrescentar, que o SWAT foi desenvolvido para simular fisicamente o comportamento hidrológico permitindo, assim, que ele seja utilizado em bacias não monitoradas (DEVIA *et al.*, 2015). Assim, há a possibilidade de investigação do impacto dos diferentes usos de solo na hidrologia, um dos objetivos da rede de monitoramento da área de estudo. E seu uso no Brasil cresceu muito nas últimas décadas (BRESSIANI, 2016).

2.6.3 SWAT

O Soil and Water Assessment Tool - SWAT é um modelo de domínio público desenvolvido conjuntamente pelo Departamento Agrícola americano

(USDA/ARS) e pelo Texas A & M AgriLife Research da Universidade Texas A & M. SWAT é um modelo físico, semi-distribuído e complexo baseado em escala de bacia hidrográfica, que funciona em um passo de tempo diário (NEITSCH *et al.*, 2011). Seu principal objetivo é a avaliação do impacto, a longo prazo, das mudanças no uso de solo e nas componentes terrestres do ciclo hidrológico. O modelo SWAT pode simular os escoamentos superficial, sub-superficial e subterrâneo, o transporte de sedimentos, nutrientes, pesticidas e bactérias da bacia modelada (ARNOLD *et al.*, 1998). Em princípio, foi desenvolvido para grandes bacias, entretanto, há diversas aplicações bem sucedidas em pequenas bacias (HARMEL *et al.*, 2006; CHO *et al.*, 2009). Inclusive, BRESSIANI (2016) faz uma revisão das aplicações do modelo SWAT no Brasil e afirma que cerca de 20% dos estudos são em bacias com área inferior a 15 km² e 72% com área inferior a 1.000 km².

Os principais componentes do modelo incluem: clima, hidrologia, propriedades do solo, características de crescimento das culturas agrícolas, nutrientes, pesticidas, bacteriologia e manejo do solo (ARNOLD *et al.*, 2012). Esses componentes são extraídos do extenso banco de dados interno do SWAT que é alimentado com características da área de estudo fornecidas pelos usuários. Assim, minimamente, os dados de entrada do SWAT correspondem a mapas de uso e ocupação do solo, tipos de solos, características topográficas da bacia (modelo digital de terreno) e dados climáticos diários (*i.e.* precipitação, temperatura máxima e mínima do ar, radiação solar, umidade relativa do ar e a velocidade do vento). Também é possível fornecer outros tipos de dados, tais com: reservatórios, transposição de vazões e usuários.

Os processos representados pelo SWAT tem como base o balanço hídrico que são simulados nas Unidades de Resposta Hidrológica (HRU) que correspondem a áreas homogêneas a partir da combinação entre uso de solo, tipo de solo e classe de declividade. Primeiramente, a área de estudo é dividida em sub-bacias de acordo com os exutórios de interesse estabelecidos pelo modelador. E, então, após o fornecimento das informações de uso e ocupação do solo, tipos de solo e classes de declividade são definidas as HRUs. O usuário pode optar em dividir a bacia hidrográfica em apenas sub-bacias, nesse caso elas serão caracterizadas por seu uso dominante de solo e tipo de solo. A simulação hidrológica da bacia é dividida em duas fases, de acordo com o local

onde ocorre a movimentação de água, sedimentos, nutrientes, pesticidas: nas sub-bacias, em direção ao canal principal, e dentro canal principal.

O escoamento superficial é estimado a partir do método Green-Ampt ou de uma versão modificada do método Curva-Número do Serviço de Conservação dos Solos americano (SCS) (padrão) agregado. O modelo possui vários métodos padrões que, em muitos casos, podem ser substituídos por dados fornecidos pelo usuário. Como por exemplo, a estimativa da evapotranspiração potencial através dos métodos de Penman Monteith (padrão), Priestly-Taylor e Hargreaves ou inserida pelo modelador. O escoamento de base é simulado através do aquífero não confinado e o fluxo lateral (subsuperficial) pelo modelo de armazenamento cinemático. A simulação dos nutrientes foi implementada com base no modelo QUAL2E (Brown and Barnwell, 1987). O modelo rastreia os nutrientes dissolvidos na corrente, que são transportados pela água, e os nutrientes adsorvidos no sedimento que serão depositados no fundo do rio. Detalhes e referências dos métodos utilizados pelo SWAT podem ser encontrados em NEITSCH *et al.*, 2009.

São disponibilizadas interfaces do modelo associadas a sistemas de informações geográficas, dentre as quais, o ArcSWAT em ArcGIS (SWAT, 2012) onde é possível realizar todo o processo de inserção e manipulação dos dados. A interface automaticamente realiza a delimitação da bacia de estudo, sub-bacias, HRUs e define a rede hidrográfica com base no MDE e definições do usuário. A rede hidrográfica pode desempenhar um papel importante nas grandes bacias hidrográficas, por outro lado, em bacias pequenas, em que o tempo de concentração é menor que um dia, ela não exerce um papel de destaque.

Após a delimitação da bacia deve-se proceder a “verificação do modelo” que consiste no processo de avaliação do código implementado e do modelo propriamente dito de forma a verificar se ele reflete o modelo conceitual (ZECKOSKI *et al.*, 2015). Nessa fase, é importante a reunião de informações sobre as características da bacia (por exemplo, valores de infiltração e evapotranspiração médios) a fim de compará-las com as saídas do modelo.

A verificação é importante, pois é possível que os dados simulados apresentem um ótimo ajuste em relação aos observados embora os valores dos parâmetros e a

configuração do modelo não represente a realidade do comportamento hidrológico. Assim ela deve ser feita durante todo o processo de modelagem. Isso é ainda mais comum com a disponibilização de software de calibração automática e uma interface amigável.

Para auxiliar nesse processo está disponível o programa "SWAT Check" (WHITE *et al.*, 2014), um programa independente do Microsoft Windows. A partir da leitura da saída do SWAT, o programa alerta o usuários de valores fora do intervalo típico referentes às componentes modeladas, quais sejam: escoamento superficial, escoamento de base, evapotranspiração, coeficientes de *runoff*, sedimentos, etc. O programa, ainda, detecta e alerta usuários de erros comuns de aplicativos modelo. Além disso, é possível visualizar a saída do modelo de forma ilustrada.

Após a realização dessa primeira verificação recomenda-se a execução da análise de sensibilidade dos parâmetros do modelo. A análise de sensibilidade é o processo através do qual se determina o impacto da mudança dos parâmetros de entrada na saída do modelo (MORIASI *et al.*, 2007). É uma ótima maneira de identificar os parâmetros mais promissores para serem usados na calibração (BRESSIANI, 2016).

De acordo com ARNOLD *et al.* (2012), há dois tipos de análise de sensibilidade: local, onde os valores dos parâmetros são alterados um de cada vez e, a global, onde os valores dos parâmetros são alterados simultaneamente. Ambas apresentam desvantagens e podem produzir resultados diferentes.

E, então, inicia-se o processo de “calibração do modelo”. A calibração consiste no ajuste dos parâmetros mais sensíveis, determinados a partir de uma análise de sensibilidade, comparando os resultados do modelo com os dados reais observados para as mesmas condições de simulação. O intuito é a redução da incerteza de predição do modelo através da sua parametrização (ARNOLD *et al.*, 2012) A calibração pode ser manual, onde os valores dos parâmetros são ajustados pelo modelador, ou automática, quando é utilizado um algoritmo, com base em uma função objetivo, para decidir quais parâmetros devem ser ajustados ou a combinação das duas denominada semi-automática (BRESSIANI, 2016; ZECKOSKI *et al.*, 2015).

Normalmente, para a modelagem de qualidade de água, inicia-se a calibração dos dados de vazão, em seguida os sedimentológicos e, por último, os de qualidade de

água. A calibração dos dados sedimentológicos pode ser desprezada em alguns casos, como por exemplo, na ausência de dados.

Em seguida, executa-se à “validação do modelo”, onde se verifica como o modelo calibrado se comporta frente a dados de entrada desconhecidos (ZECKOSKI *et al.*, 2015). Nesse caso, não há alteração dos parâmetros do modelo ajustados durante o processo de calibração, apenas dos dados de entrada, que devem ser diferentes daqueles usados na calibração.

Para a realização dos processos de calibração e validação, tradicionalmente, há a divisão dos dados observados em dois conjuntos: um para ser usado durante a calibração e outro para ser usado na validação, embora existam outras técnicas. Normalmente, essa divisão é feita com base na série temporal, ou seja, são obtidos dois conjuntos de dados, em períodos diferentes, para o(s) mesmo(s) ponto(s) de monitoramento. Entretanto, os dados também podem ser espacialmente divididos. Nesse caso, toda a série de dados de uma estação é usada para calibração enquanto os dados da(s) outra(s) estação (ões) são utilizados para a validação. A validação espacial é utilizada em casos especiais quando, por exemplo, há poucos dados disponíveis (ARNOLD *et al.*, 2012).

O software SWAT-CUP (ABBASPOUR *et al.*, 2007) é um programa de domínio público que foi desenvolvido para auxiliar nos processos de análise de sensibilidade, calibração, validação e análise de incerteza do modelo SWAT. São disponibilizados diferentes processos de calibração. O SUFI2 é um dos procedimentos mais usados dentre as aplicações do SWAT e permite a calibração manual ou automática dos parâmetros que é avaliada através de diversas métricas estatísticas (BRESSIANI, 2016).

O SWAT possui algumas limitações que devem ser destacadas. O SWAT pode se tornar uma ferramenta com uma entrada de dados altamente intensiva. Pois, embora o SWAT seja desenvolvido para usar dados prontamente disponíveis isso nem sempre é o caso, especialmente nos países em desenvolvimento (CAMBIEN, 2017). Além disso, a precisão dos dados de entrada pode ser um problema já que pode prejudicar os dados de saída com avaliação complicada (ARNOLD *et al.*, 2012). Além disso, alguns dos processos físicos são simplificados sem que haja uma forma de avaliar o impacto dessas simplificações (TETRA TECH, 2004).

O Swat é um dos modelos mais utilizados no mundo (BRESSIANI, 2016; CAMBIEN, 2017). No Brasil, o uso do SWAT tem aumentado nos últimos anos, com aplicações em diversas áreas de tal forma que a localidade de Porto de Galinhas, em Pernambuco (RIBEIRO *et al.* 2015, BRESSIANI, 2016).

3 Caracterização da Área de Estudo

Nessa seção será apresentada uma caracterização geral da Bacia do Rio Piabanha onde foram aplicadas as metodologias desenvolvidas nessa tese. A seleção dessa bacia ocorreu em razão dos diversos problemas ambientais que acometem a região, os quais requerem um monitoramento hidrometeorológico intensivo para que sejam estudados, compreendidos e solucionados, conforme será exposto a seguir.

3.1 Caracterização Física

A área de estudo está localizada na região serrana do estado do Rio de Janeiro, na bacia hidrográfica do rio Piabanha. O rio Piabanha é afluente pela margem direita do rio Paraíba do Sul, uma bacia de relevância nacional por abrigar municípios dos estados do Rio de Janeiro, São Paulo e Minas Gerais. A bacia ocupa uma área de aproximadamente 2.050 km² e está totalmente inserida no estado do Rio de Janeiro sendo composta pelos municípios de Areal, Paraíba do Sul, Paty do Alferes, Petrópolis, São José do Vale do Rio Preto, Teresópolis e Três Rios (ARAÚJO *et al.*, 2007) conforme a Figura 10.

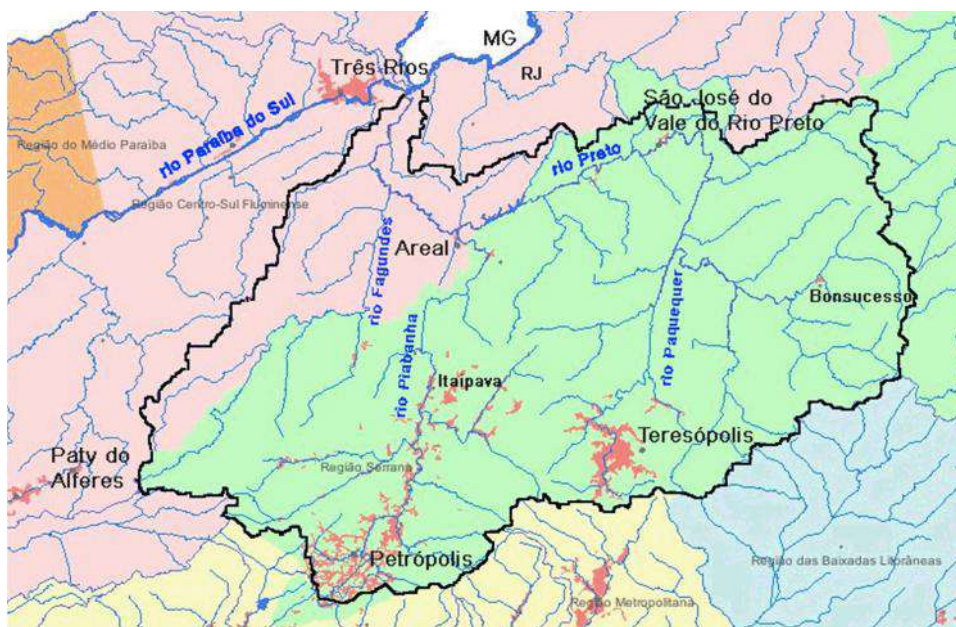


Figura 10 - Localização da Bacia do rio Piabanha no mapa de regiões político-administrativas do Estado do Rio de Janeiro (Fonte: Fundação CIDE, LabGeo in FUNDAÇÃO COPPETEC, 2010).

A bacia apresenta relevo montanhoso, de modo geral, muito acidentado, no curso médio e superior, com afloramentos rochosos e altitudes que ultrapassam os 2.000m. A maior parte dos cursos d'água é do tipo encaixado com grandes declividades fazendo com que os níveis se elevem e abaxem rapidamente em resposta a um evento chuvoso. Assim, para registrar eventos de cheia, ou seja, as vazões máximas é necessário um monitoramento contínuo com frequência sub-diária (FUNDAÇÃO COPPETEC, 2010). A Figura 11 apresenta o perfil longitudinal do Rio Piabanha até a estação Pedro o Rio onde é possível notar a alta declividade ao longo de toda a extensão do rio com poucas zonas efetivamente planas (ARAÚJO, 2016).

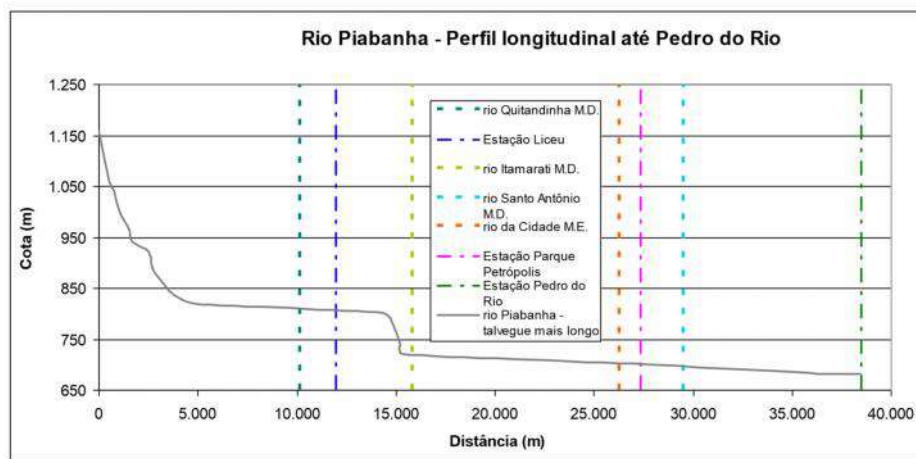


Figura 11. Perfil longitudinal do Rio Piabanha (Fonte: ARAÚJO, 2016).

A região apresenta problemas graves de deslizamentos de encostas, proporcionados, em parte, pela ocupação irregular nas áreas de alta declividade (grande parte da área da bacia). Pois, devido à proximidade com a cidade do Rio de Janeiro o processo de desenvolvimento foi acentuado e inadequado para as suas condições ambientais. ARAÚJO (2016) reuniu dados do Universo do Censo Demográfico 2010, para compor a distribuição demográfica da bacia, juntamente com os lançamentos e captações cadastradas no Cadastro Nacional de Usuários de Recursos Hídricos (CNARH-2012) no mapa apresentado na Figura 12. É possível observar grandes concentrações de moradores (áreas vermelhas do mapa) próximas aos limites da bacia que são regiões de altas declividades, evidenciando o problema anteriormente citado de deslizamentos de terra por ocupação de encostas. O que é observado, também, na carta de suscetibilidade a movimentos gravitacionais de massa e inundação do município de Petrópolis para a área da bacia representativa apresentada na Figura 13 (CPRM, 2013).

Nota-se cicatrizes de deslizamento bem como corrida de massa em áreas com alta concentração de moradores. Outro aspecto peculiar é a grande ocupação urbana em áreas de conservação, evidenciando a carência de um planejamento urbano adequado.

Além disso, a bacia sofre constantemente com inundações, conforme áreas de enxurradas e inundação observadas na figura 13. A qualidade da água dos rios é deteriorada principalmente pelo lançamento de esgotos. O tratamento dos esgotos é precário e as vazões naturais dos rios são muito baixas para a diluição das cargas lançadas, resultando em índices consideráveis de poluição, principalmente, na parte urbana (ARAÚJO *et al.*, 2007). A cidade de Petrópolis tem avançado muito nessa questão, principalmente após a privatização do setor. Atualmente, segundo dados da concessionária “Águas do Imperador”, o sistema de esgoto da cidade é composto por quatro estações de tratamento principais, além de outras 12 unidades de tratamento de esgoto em bairros, dez biodigestores e um biossistema, que juntos tratam 83% dos esgotos urbanos. Dessa forma é possível que nos próximos anos haja uma melhoria da qualidade de água nessa área da bacia.

Na Figura 12 observa-se a delimitação de uma sub-bacia da bacia do Rio Piabanha denominada “bacia representativa” que corresponde à área de estudo selecionada para esse trabalho conforme será detalhado na seção 3.2.3. Nota-se que pouquíssimas captações e lançamentos cadastrados estão localizados na área de estudo. Sendo assim, num primeiro momento, esses não serão considerados na aplicação da metodologia.

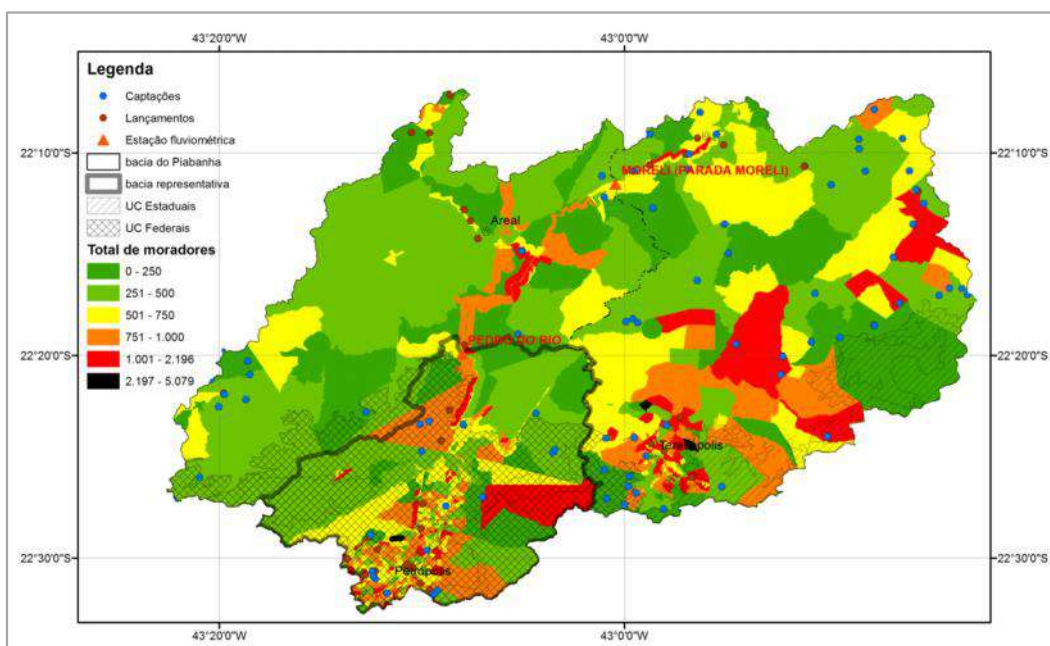


Figura 12 - Distribuição da população por setores censitários com captações de água e lançamentos de efluentes (ARAÚJO, 2016).

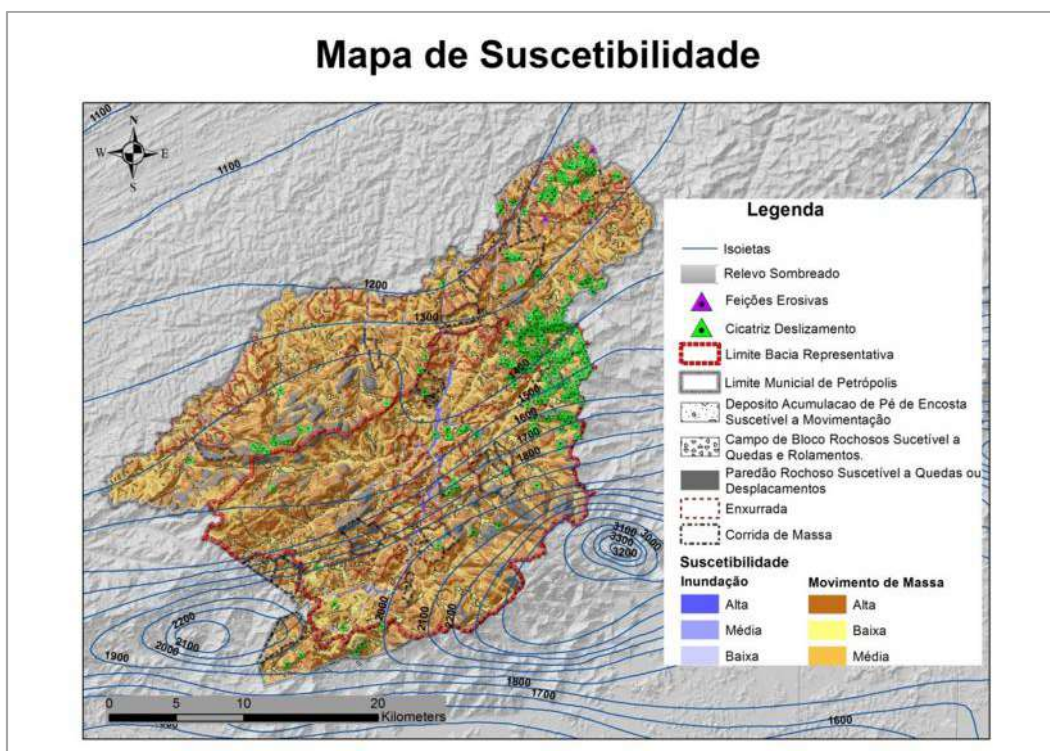


Figura 13 – Carta de suscetibilidade a movimentos gravitacionais de massa e inundações do município para a bacia representativa associada às Isoietas Médias Anuais (adaptado de CPRM/2010 e CPRM/2013).

3.2 As bacias experimentais e representativas e o Projeto

EIBEX

A área selecionada para esse estudo compreende uma sub-bacia da bacia do Rio Piabanha que é monitorada pelo Serviço Geológico do Brasil – CPRM através do projeto Estudos Integrados em Bacias Experimentais e Representativas – Região Serrana/RJ – EIBEX desde o ano de 2007.

3.2.1 Definições

Bacias representativas são aquelas consideradas representativas de uma região hidrológica e são utilizadas para investigações intensivas de problemas específicos do ciclo hidrológico. Recomenda-se que o monitoramento seja de longo termo e, se possível, combinado com o estudo das características climatológicas, pedológicas, geológicas e hidrogeológicas (TOEBES *et* OURYVAEV, 1970). Elas devem representar a realidade sócio, econômica, física e ambiental, possibilitando a extrapolação dos resultados dos experimentos para a bacia principal ou áreas similares (Pimentel da Silva *et al.*, 2010). De acordo com PAIVA *et* PAIVA (2001), são sub-bacias instrumentadas com aparelhos de observação e registro de fenômenos hidrológicos que representam bacias situadas em uma mesma região homogênea cuja observação deve ser realizada por longos períodos de tempo, preferencialmente superiores a 30 anos.

Bacias experimentais são bacias relativamente homogêneas no que se refere à cobertura do solo e vegetação, possuem características físicas uniformes e são deliberadamente modificadas para o estudo detalhado do ciclo hidrológico (TOEBES *et* OURYVAEV, 1970). Num primeiro momento, as bacias experimentais eram essencialmente bacias preservadas que mantinham suas condições naturais. Entretanto, com o tempo verificou-se a necessidade, em alguns casos, de selecionar bacias com algumas condições naturais alteradas para estudar o efeito das modificações sobre o comportamento hidrológico, inferindo leis e demais relações (PAIVA *et* PAIVA, 2001). Elas funcionam como pequenos laboratórios, densamente equipados, estabelecidos no mundo real que contribuem para a compreensão das relações e processos envolvidos no

ciclo hidrológico fornecendo suporte, também, para o treinamento e capacitação (Šraj *et al.*, 2008; MEDEIROS *et al.*, 2005 in PIMENTEL DA SILVA *et al.*, 2010). As bacias experimentais destacam-se, ainda, no contexto da regionalização das informações, como meio de caracterizar com maior precisão as relações entre solo, água, vegetação e atmosfera e transpor esse conhecimento, juntamente com a modelagem, para regiões “hidroclimatologicamente” semelhantes sem monitoramento dessas relações (RODRIGUES, 2014).

3.2.2 Contexto Histórico

De acordo com GU *et al.* (2013) os primeiros estudos modernos em bacias surgiram no final do século XIX na Suíça, sendo sucedidos por uma multidão de estudos até meados do século passado em diversas partes do mundo, tais como (países e ano de início dos estudos em bacias experimentais e representativas): Estados Unidos (1910), Rússia (ex-União Soviética) (1933), Alemanha (1948), França (1950) e China (1953).

Em seguida, houve um período de rápido desenvolvimento das pesquisas na área, resultante, principalmente, da “Década de Hidrologia Internacional” (International Hydrology Decade – IHD), entre os anos de 1965-1974 (NACE, 1969). A IHD foi lançada pela Conferência Geral da UNESCO em sua décima terceira sessão para promover a cooperação e intercâmbio internacional em pesquisa, estudos e treinamento de especialistas e técnicos em hidrologia científica (TOEBES *et OURYVAEV*, 1970). O programa da IHD incluía um tópico específico sobre Bacias Representativas e Experimentais que deveria ser aprofundado e detalhado.

Dessa forma, em 1965, a UNESCO organizou um Simpósio sobre Bacias Representativas e Experimentais, na Polônia, a fim de estudar as experiências e resultados obtidos por diversos países na área. Como resultado elaborou e publicou um guia para pesquisa e prática internacional a fim de definir princípios metodológicos e de funcionamento para as Bacias Experimentais e Representativas (TOEBES *et OURYVAEV*, 1970).

Por outro lado, o grande crescimento e disseminação da pesquisa também gerou polêmica que culminou com a publicação de alguns artigos com críticas duras ao tema. As críticas às Bacias Experimentais e Representativas incluíam diversos aspectos, tais como: os altos custos de manutenção envolvidos, a falta de progresso no conhecimento básico hidrológico, carência de representatividade e dificuldade de transferência dos resultados das pequenas para grandes áreas (HEWLETT *et al.*, 1969; LEOPOLD, 1970). A demasiada espera por resultados também era constantemente debatida, já que a construção de séries de dados confiáveis leva tempo e envolve problemas dos mais diferentes níveis, desde o bom funcionamento de equipamentos até o armazenamento dos dados de forma eficiente.

Embora tais críticas tenham levado a reflexão, principalmente, da relação “custo-benefício” das bacias experimentais e representativas, elas não foram suficientes para abalar o avanço nas pesquisas desenvolvidas. Tanto é assim que, nas últimas décadas, observa-se a consolidação desses experimentos e, principalmente, a formação de redes de instituições que visam à troca de experiência, fortalecimento e manutenção das pesquisas. Como, por exemplo, a rede de bacias experimentais implantada na década de 30 pela divisão de hidrologia do USDA-ARS, operada ou financiada pelo USGS, USDA Forest Service, USDI-NPS e NSF, com série de dados entre 38 e 71 anos de duração. Os estudos desenvolvidos produziram, e continuam produzindo, a base científica para vários modelos atualmente usados no mundo inteiro. A série de dados hidrológicos de longo termo subsidiou a compreensão e gerenciamento dos recursos hídricos em bacias de diferentes regiões (HARMEL *et al.*, 2007).

Outro exemplo é a Rede Euromediterrânea de Bacias Experimentais e Representativas (ERB) fundada em 1986 tendo como principais objetivos reunir cientistas europeus envolvidos na pesquisa de hidrologia em bacias e estimular o intercâmbio de informação e resultados, bem como encorajar a cooperação em projetos de pesquisa. A ERB reúne, atualmente, 23 países. Ela realiza conferências bianuais, a publicação de um boletim semestral e a manutenção de um registro de bacias hidrográficas para estudos de longo termo sobre o balanço hídrico e pesquisa em áreas afins. Grande parte da pesquisa refere-se a pequenas bacias naturais, mas o desenvolvimento de métodos para medir, monitorar e avaliar os impactos das mudanças

ambientais também faz parte de vários de seus estudos (ERB, 2017). A ERB recebe apoio financeiro contínuo da UNESCO e dos países envolvidos.

No Brasil, as primeiras bacias experimentais foram instaladas na região nordeste, na década de 70 em razão da necessidade da melhor compreensão do comportamento hidrológico da região em função das condições climáticas desfavoráveis marcadas pelos frequentes períodos de secas (ALVERGA, 2016). As primeiras bacias experimentais foram implantadas por meio de uma parceria entre a Superintendência de Desenvolvimento do Nordeste (SUDENE), o Instituto Francês de Pesquisa Científica para o Desenvolvimento em Cooperação (ORSTOM) e universidades locais (ALVERGA, 2016).

Em 2001, nessa mesma linha, através da iniciativa de um grupo de professores, estudantes e pesquisadores das seguintes instituições: UFBA, UFPE, UFRPE, UFPB, UFCG, UFRN, UFAL, UFC e FUNCEME (Fundação Cearense de Meteorologia e Recursos Hídricos) foi criada a Rede Hidrológica do Semiárido (REHISA) (MONTENEGRO *et al*, 2012). Através dessa iniciativa, com financiamentos de órgãos de fomento como FINEP e CNPq, em 2002, desenvolveu-se o projeto de Implantação das Bacias Experimentais do Semiárido (IBESA), responsável pela implantação de sete Bacias Experimentais no Semiárido nordestino e, em 2005, Projeto BEER responsável pelo início da instrumentação das bacias.

Como citado anteriormente, o suporte dos órgãos de fomento é essencial para a consolidação da pesquisa no Brasil. Assim, pode-se destacar, a Chamada Pública MCT/FINEP CT-HIDRO 01/2010 que teve como um dos principais objetivos a integração das instituições de pesquisa através da formação de redes em torno dos temas prioritários como as Bacias Representativas Urbanas e Rurais (FINEP, 2010). Um dos resultados foi a criação, ano de 2011, da Rede REHIDRO (Rede Nacional de Bacias Experimentais – Finep/CT-Hidro) composta pelas universidades: UFPE, UFRPE, UFAL, UnB/EMBRAPA CERRADOS e UFS/EMBRAPA TABULEIROS COSTEIROS (MONTENEGRO *et al.*, 2014). . A atuação principal dessa rede é a pesquisa com o intuito da compreensão do comportamento hidrológico nos Biomas Caatinga e Cerrado (MONTENEGRO *et al.*, 2015).

Dessa forma, é possível observar, no Brasil, iniciativas espaço-temporal isoladas, por meio de universidades financiadas, principalmente, por órgãos públicos de fomento, carecendo de um apoio institucional público contínuo. Esse suporte é muito importante para manutenção dos experimentos já que muitos precisam de um longo tempo para apresentar resultados. Entretanto, o desafio está na continuidade dos estudos pois nota-se um incentivo inicial para a implantação das bacias que se perde a longo do tempo por falta de comprometimento institucional. Em contrapartida, observa-se nas iniciativas americana e européia um envolvimento institucional público, principalmente, de órgãos de fomento, ao longo de todo o processo, contribuindo para obtenção de resultados contínuos.

Atualmente, a CPRM realiza estudos em bacias experimentais e representativas em diversos biomas brasileiros (i.e. Cerrado, semi-árido e mata atlântica) em parceria com universidades, órgãos gestores e, muitas vezes, com apoio de agências governamentais de fomento. O objetivo é estabelecer um modelo de parcerias institucionais que possa ser replicado e contribua para a consolidação dessa área de pesquisa. Um exemplo é o projeto EIBEX, base para o estudo desenvolvido nessa tese.

3.2.3 O projeto EIBEX

O projeto EIBEX tem como objetivo a avaliação do comportamento hidrológico em região com bioma natural Mata Atlântica e em áreas de ocupação agrícola e urbanizada, que são os principais usos de solo da bacia do Piabanha. O projeto visa desenvolver pesquisa e estudos na área de hidrologia, com ênfase em: climatologia, qualidade da água, solos, GIS, e também o uso de diferentes tecnologias de medição de dados com base em uma rede de monitoramento hidrometeorológica (VILLAS-BOAS *et al.*, 2017).

Essa rede foi estabelecida em uma sub-bacia da bacia do rio Piabanha, com cerca de 400km², que se estende desde a nascente do Rio Piabanha até a estação fluviométrica Pedro do Rio (código 58405000), que compõe a Rede Hidrometeorológica Nacional e possui uma série longa de dados consistentes. Ela reúne as características de uso de solo e vegetação da bacia do rio Piabanha e, assim, foi definida como *bacia representativa* de forma a extrapolar os resultados obtidos para a bacia principal. Dentro desta área,

foram definidas três bacias *experimentais*, onde predominam os diferentes usos do solo existentes: em área de mata Atlântica preservada (Figura 14), em área predominantemente de uso agrícola (Figura 15) e área de ocupação urbana (Figura 16), respectivamente, com as seguintes áreas: 47 km², 30 km² e 13km². A Figura 17 apresenta a localização da bacia do rio Piabanha e as bacias representativa. A partir do mapa de uso e ocupação da bacia (escala 1:25.000) (INEA, 2017), apresentado na Figura 30, nota-se que a maior parte da área da bacia do Piabanha é ainda ocupada por florestas, vegetação e áreas naturais (i.e. afloramentos rochosos). A área agrícola é pouco expressiva quando comparada com essas classes, mas é influente quando comparada com a urbana, destacando-se a agricultura dedicada à produção de olerícolas. Os detalhes da seleção das bacias e instalação das estações integrantes do projeto podem ser encontrados em Mascarenhas (2007).



Figura 14 - Paisagem a montante da estação Rocío, no município de Petrópolis, na bacia de Mata Preservada.



Figura 15 - Área de cultivo na encosta e uma parte do relevo bem acentuado, região do Bonfim, na bacia de uso Agrícola.



Figura 16 - Estação Liceu no Centro de Petrópolis, na bacia de uso Urbano.

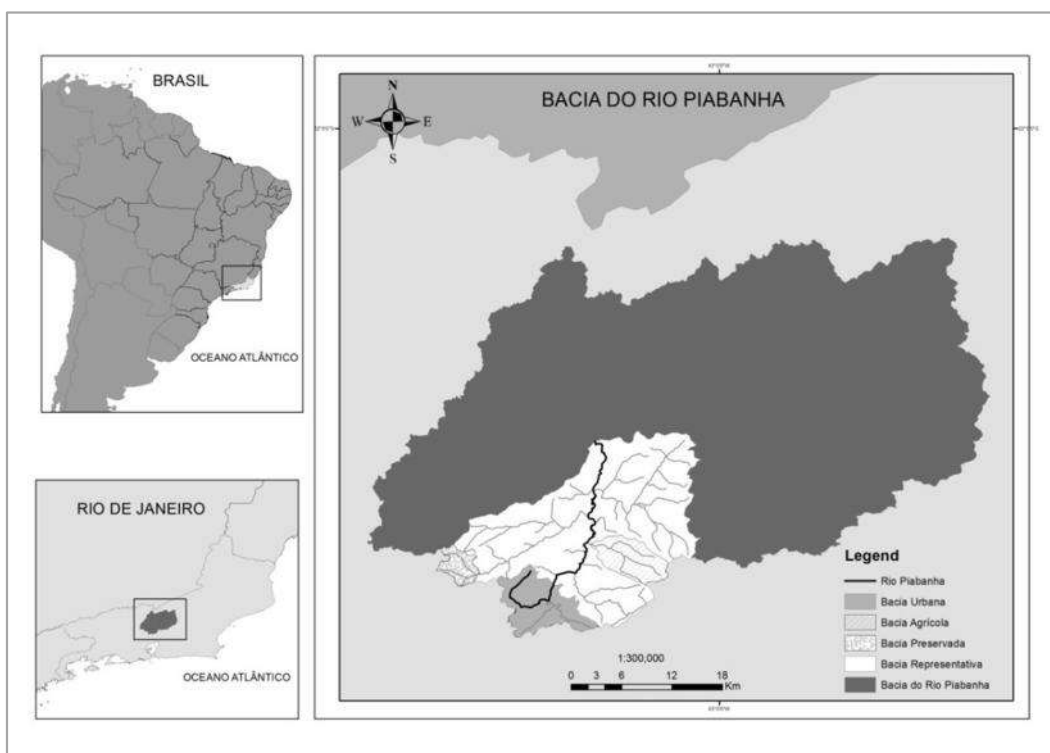


Figura 17 - Localização da bacia do rio Piabanha e das bacias representativa e experimentais. (adaptado de VILLAS-BOAS *et al.*, 2017)

O projeto é apoiado por três "pilares": integração com instituições que desenvolvem estudos na bacia, teste de equipamentos e desenvolvimento de pesquisas na área de hidrologia. A integração interinstitucional visa a troca de experiência e conhecimento entre a CPRM e as diversas instituições atuantes na área da bacia. Nesses 10 anos de projeto a CPRM participou de diversos projetos de cooperação interinstitucional que fomentaram o desenvolvimento dos experimentos. Entretanto, pode-se observar que o principal responsável pela continuidade dos estudos foi o fato do projeto ter um orçamento público anual, através CPRM, para operação e manutenção da rede de monitoramento. A equipe do projeto está sempre em busca de novas tecnologias e equipamentos para serem implementadas na bacia com intuito de realizar testes nos equipamentos e capacitação da equipe. Por último, o projeto tem o intuito de desenvolver pesquisa e estudos com base nos dados oriundos da rede de monitoramento hidrometeorológica instalada na bacia e operada pela CPRM (VILLAS-BOAS *et al.*, 2017).

A rede de monitoramento hidrometeorológico do projeto EIBEX (RMHE), atualmente, conta com 13 estações conforme apresentado na Tabela 1 e na Figura 18.

Atualmente as estações contam com medições convencionais (pluviômetro (P) e régua limnimétrica (P)) e com equipamento automático (r). Há duas estações climatológicas na bacia (C) que fazem as seguintes medições: pressão atmosférica, umidade relativa do ar, temperatura do ponto de orvalho, velocidade e direção do vento, radiação solar incidente, radiação líquida, evaporação, potencial matricial, fluxo de calor e umidade do solo, temperatura do solo e precipitação. Há 5 estações que são telemétricas (T). Em todas as estações fluviométricas são realizadas medições de vazão (D) a cada dois meses e medições de qualidade de água (Q) com frequência variável de acordo com o estudo. Os dados de monitoramento ainda não estão sendo disponibilizados *on line* mas podem ser requeridos através do website da CPRM. A operação da maior parte das estações teve início em 2007 com medições quantitativas, e em 2009 com qualidade da água.

Tabela 1 – Informações das estações de monitoramento do projeto EIBEX.

ESTAÇÃO	Sigla	Código FLU	Código PLU	TIPO	CURSO D'AGUA	LATITUDE	LONGITUDE	Data de início da operação - PLU/FLU	Data de início da operação - QA
Pq.Petropolis	PP	58400250	2243286	CFDFrQT	Rio Piabanha	22° 24' 19"	43° 08' 00"	1/8/09	27/8/09
Esperança	E	58400010	2243287	PPRFDFrQ	Rio Piabanha	22° 30' 39"	43° 12' 37"	28/4/07	27/8/09
Liceu	L	58400050	2243289	PPRFDFrQT	Rio Piabanha	22° 29' 14"	43° 10' 38"	24/4/07	27/8/09
Morin	M	58400030	2243288	PPRFDFrQ	Rio Palatinado	22° 31' 00"	43° 10' 08"	22/4/07	27/8/09
Poço Tarzan	PT	58400110	2243303	PPRFDFrQT	Rio Bonfim	22° 27' 14"	43° 06' 28"	23/4/07	27/8/09
Poço do Casinho	PC	58400104	****	FDfrQ	Rio Açú	22° 27' 39,6"	43° 05' 40,8"	31/10/07	27/8/09
Joao Christ	JC	58400108	****	FDQ	Rio Alcobaça	22° 27' 37,19"	43° 05' 59,76"	28/10/07	27/8/09
Pedro do Rio	PR	58405000	*****	PPRFDFrQT	Rio Piabanha	22° 19' 56"	43° 08' 01"	1/8/30	27/8/09
Vila Açú	VA	****	2243301	PPR	****	22° 27' 45,20"	43° 05' 29,30"	1/11/09	-
Sítio das Nascentes	SN	****	2243291	C	****	22° 28' 7,63"	43° 06' 9,21"	31/10/2007e1 9/4/2016*	-
Rocio 2 - Ponte	R	58400212	****	FDQ	Rio da Cidade	22° 28' 38,70"	43° 15' 24,60"	28/4/10	27/8/09
Rocio 2 - D	RD	58400210	2243302	FDfr	Rio da Cidade	22° 28' 38,86"	43° 15' 28,95"	1/4/10	-
Rocio 2 - E	RE	58400211	*****	PPRFDFrT	Rio da Cidade	22° 28' 37,69"	43° 15' 27,66"	1/4/10	-

*A estação começou apenas pluviométrica em 2007 e, em 2016, foi transformada em climatológica

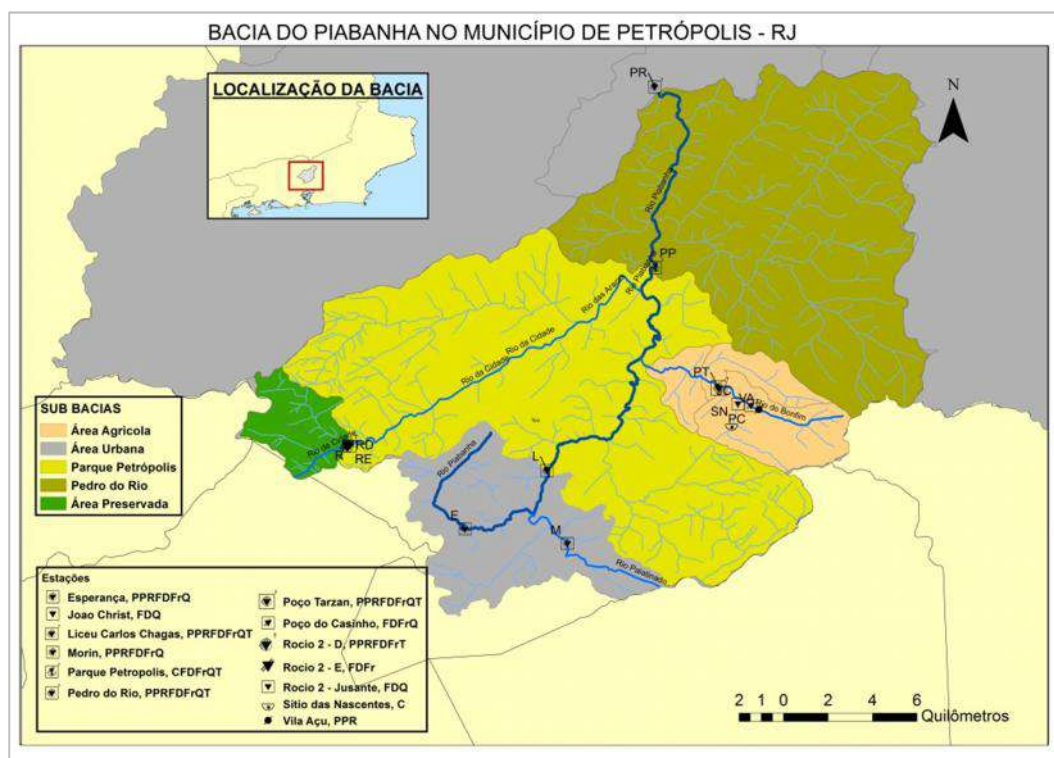


Figura 18 – A Rede de Monitoramento Hidrometeorológica do projeto EIBEX (VILLAS-BOAS *et al.*, 2017).

3.3 Clima e regime hidrológico

O ano hidrológico da bacia compreende o período de setembro a agosto, com período chuvoso entre os meses de setembro e abril (CPRM, 2017). A distribuição das precipitações ao longo do ano identifica JANEIRO, como mês mais úmido, e JULHO, como mês mais seco (ANDRADE, 2016). Nas encostas íngremes, a pluviosidade média anual ultrapassa os 2.000 mm, como nas cidades de Petrópolis e Teresópolis. e nas proximidades dos municípios de Areal e São José do Vale do Rio Preto, a média pluviométrica decresce abruptamente para 1.300 mm, com períodos secos e déficits hídricos bastante pronunciados (GONÇALVES, 2008; LOU, 2010; DE PAULA, 2011; ANDRADE, 2016; ARAÚJO, 2016). Na Figura 19 é possível observar a distribuição anual das precipitações para a bacia representativa com base nos totais mensais médios das estações com séries mais longas disponíveis (ARAÚJO, 2016).

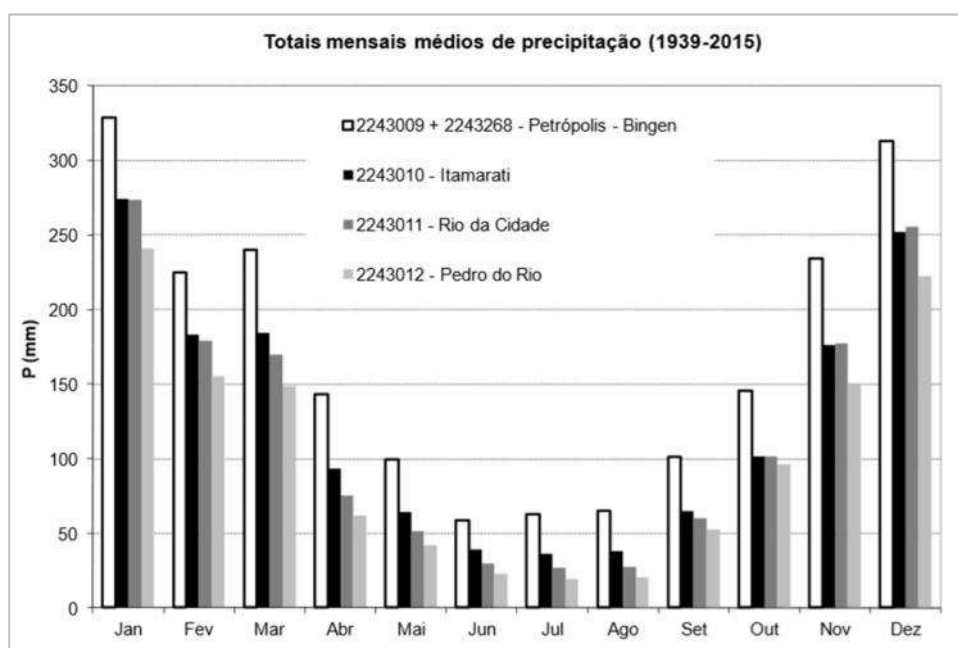


Figura 19 - Regime de precipitações na bacia representativa do rio Piabanha (Fonte: ARAÚJO, 2016).

ARAÚJO (2016) faz uma extensa análise hidroclimatológica da bacia do Piabanha e afirma que devido ao seu relevo acentuado há uma grande variação nos aspectos meteorológicos quando comparada a sua pequena área. A autora pontua que fenômenos climáticos como: a Zona de Convergência do Atlântico Sul (ZCAS), a Zona de Convergência de Umidade (ZCOU), El Niño Oscilação Sul (ENSO) e a Oscilação Multidecadal do Atlântico – (OMA ou AMO) exercem influência sobre a variação pluviométrica da região e faz uma análise detalhada das séries históricas de precipitação em relação a esses fenômenos. Conclui que a variabilidade temporal das chuvas na bacia representativa tem um comportamento sazonal que é perturbado em anos de episódio de *El Niño* e que a variabilidade espacial é caracterizada pelo relevo e pela orientação dos sistemas atmosféricos. Enfatiza, ainda, a importância do estabelecimento de uma rede de monitoramento bem distribuída para uma boa representação de ambas variabilidades. A Figura 20 mostra o mapa da bacia representativa com a rede de monitoramento onde se observa as isoietas anuais médias (1997-2006) (CPRM, 2010) sobre o modelo digital de terreno extraído do SRTM 30m. É possível notar a distribuição espacial da chuva que varia no intervalo de 1300-2500 mm/ano. Um núcleo com valores elevados de precipitação próximos à cabeceira da bacia agrícola onde estão localizadas as maiores altitudes.

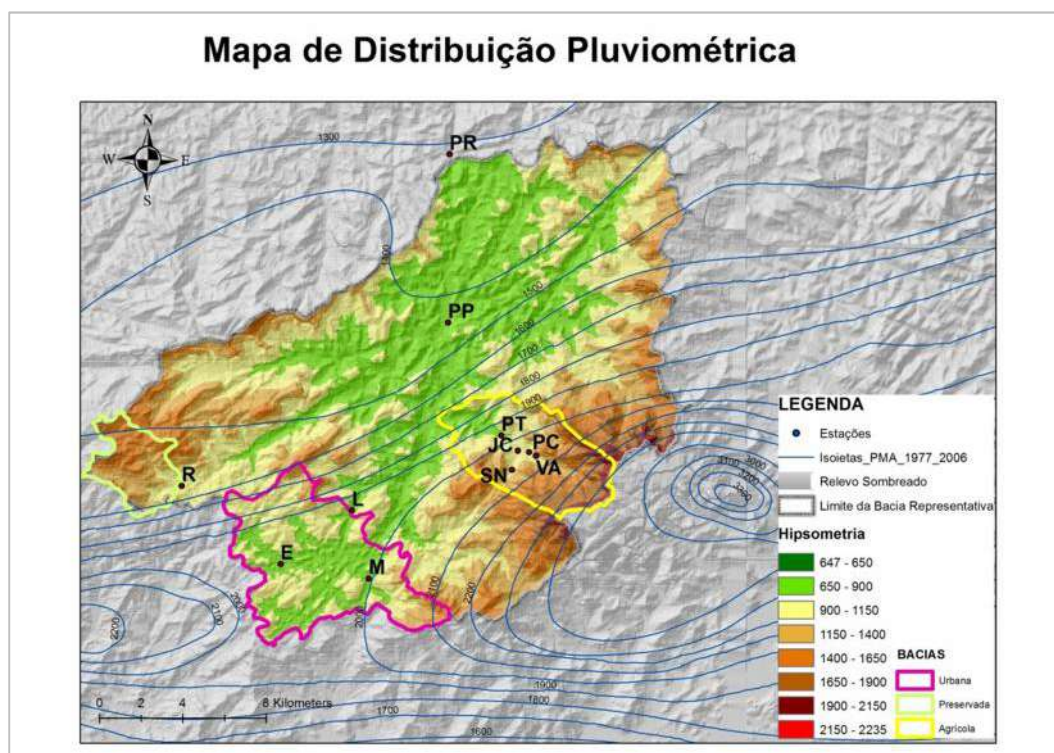


Figura 20. Mapa de isoietas da bacia hidrográfica sobre o modelo digital de terreno baseado no SRTM 30m (adaptado de CPRM-2010).

A Figura 21 apresenta os totais anuais para o período de 2011-2015 das estações convencionais das estações da RMHE: Parque Petrópolis, Rocio 2D, Sítio das Nascentes e Vila Açú e Liceu. Vale ressaltar, que alguns anos apresentaram falhas e, só foram utilizados os totais anuais de precipitação com falhas em no máximo dois meses secos ou um mês úmido. É possível analisar aqui, também, a variabilidade espacial da precipitação na bacia representativa. Observa-se que a estação Rocio apresenta os maiores valores de precipitação total anual. Essa estação fica localizada na bacia preservada em uma altitude elevada (Figura 20) e na maior parte dos anos acima de 2000mm. Por outro lado, a estação Parque Petrópolis apresentou os menores totais anuais. Ela fica localizada na parte mais baixa da bacia representativa, próxima a exutório. As estações Sítio das Nascentes e Vila Açú ficam localizadas na área agrícola, mas a primeira fica numa altitude maior justificando seus maiores índices pluviométricos em comparação a segunda. A Figura 21 apresenta uma característica muito interessante quando comparamos os totais anuais de precipitação ao longo dos anos. Observa-se que os anos 2014 e 2015 apresentam uma precipitação bem inferior ao

demais, o que se justifica pela grande estiagem enfrentada pela bacia entre os anos de 2013 e 2015 (ARAÚJO, 2016).

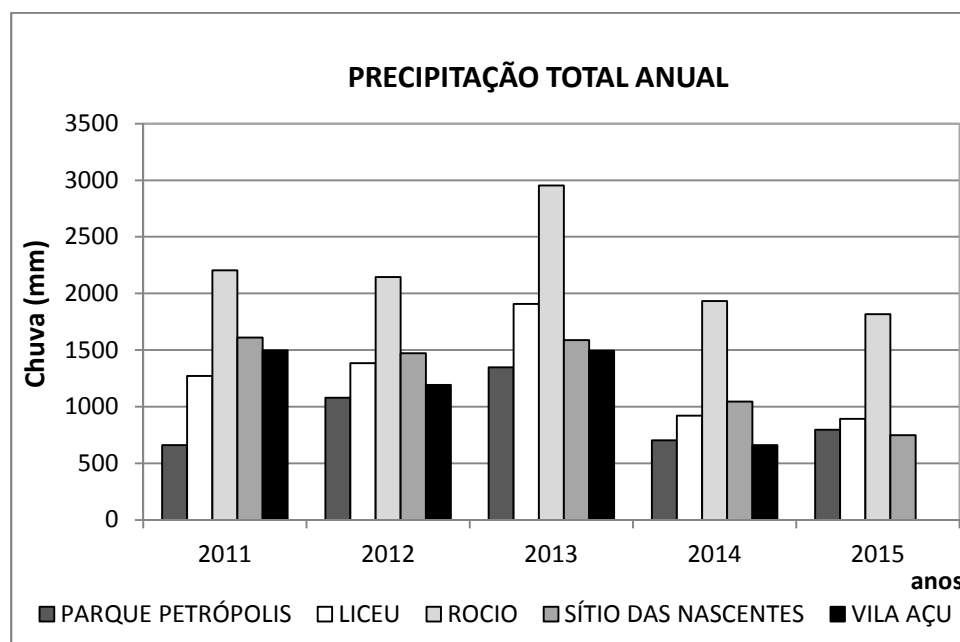


Figura 21 - Totais anuais de precipitação das estações Parque Petrópolis, Liceu, Rocio, Sítio das Nascentes e Vila Açú (adaptado de SILVA *et al.*, 2017) .

O regime de vazões da bacia do Piabanha corresponde ao ano hidrológico onde se observam as maiores vazões entre os meses de dezembro e março e, as mínimas, entre julho e setembro, conforme histograma de vazões apresentado na Figura. Como já comentado na seção 3.1, devido às características fisiográficas da bacia, o tempo de concentração é muito pequeno, fazendo com que as vazões máximas apresentem uma permanência muito baixa também. GONÇALVES (2008) e ARAÚJO (2016) estimaram o tempo de concentração até a estação Pedro do Rio correspondente a oito horas. Com base nos dados dessa estação, ARAÚJO (2016) ainda calculou a permanência da vazão média de cheia abaixo de 0,5% e afirmou que existe uma alta variabilidade das vazões ao longo do ano, como pode ser observado na Figura que apresenta série histórica de vazões diárias no período de 1931-2015. Essas características da bacia enfatizam a necessidade um planejamento da rede de monitoramento de forma a capturar tais peculiaridades.

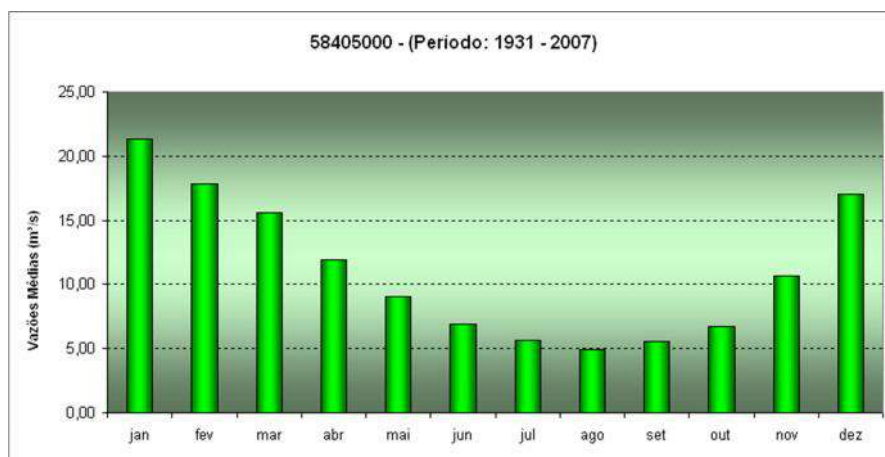


Figura 22- Vazões médias mensais do rio Piabanha na estação Pedro do Rio (58405000) (CPRM, 2017).

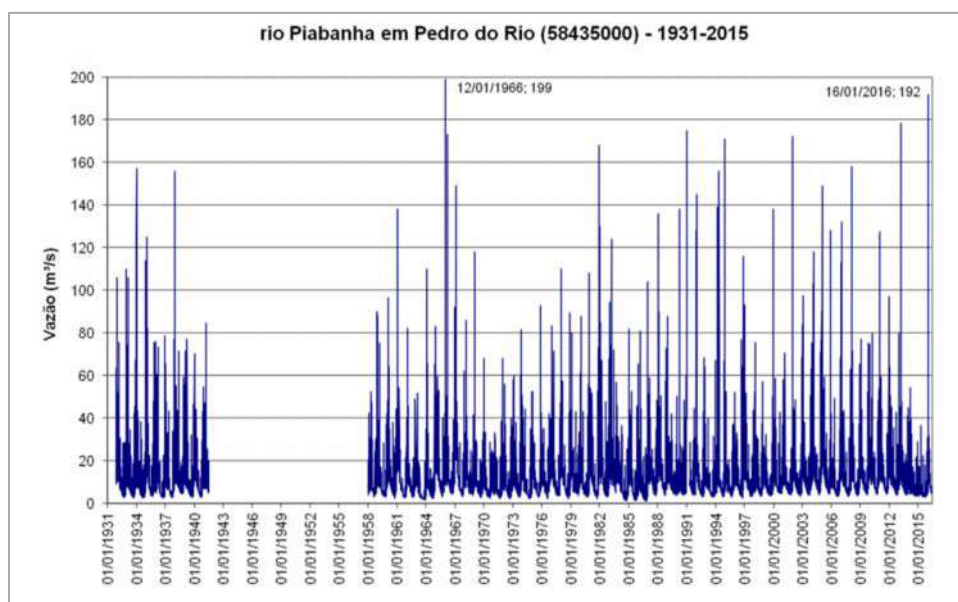


Figura 23. Vazões diárias com base nos dados observados na estação Pedro do Rio no período de 1931-2015 (Fonte: ARAÚJO, 2016).

No que concerne a evapotranspiração na bacia, ainda há poucos estudos que aprofundam o tema. GONÇALVES (2008) estimou a evapotranspiração, para o período de 1998-2004, por meio da metodologia do balanço hídrico sazonal que leva em conta os dados de chuva e vazão e oferece uma alternativa ao cálculo da evapotranspiração pelos métodos empíricos tradicionais que necessitam de variáveis meteorológicas. O autor encontrou um valor médio anual de 766 mm/ano, constatando uma ligeira tendência de superestimativa da evapotranspiração, contudo, afirma que os resultados

podem ser considerados aceitáveis dentro da margem de erros admitida na modelagem hidrológica. BARROS (2012) utilizou o modelo SEBAL e imagens Landsat em conjunto com dados reais de estações meteorológicas, incluindo a estação Parque Petrópolis, da RMHE, e formulações empíricas para estimar valores de evapotranspiração para a bacia do Rio Piabanha. Segundo os resultados encontrados, o modelo SEBAL superestimou a evapotranspiração real para a área da estação meteorológica Parque Petrópolis com valores variando entre 1,99 a 2,53mm/dia, o que significa uma evapotranspiração anual variando de 726 a 923 mm/ano aproximadamente. ARAÚJO (2016) calculou a evapotranspiração potencial usando os dados da estação Parque Petrópolis e a formulação empírica de Penain-Monteih e encontrou o valor de 959mm. A autora apresenta alguns resultados de outros estudos que estimaram a evapotranspiração real e a potencial para regiões de Mata Atlântica, como o caso de DE CICCIO (2009) que encontrou os valores totais anuais médios de evapotranspiração potencial da ordem de 1007 mm e real 738 mm.

3.4 Cartografia selecionada para o estudo

A bacia do rio Piabanha possui uma disponibilidade cartográfica razoável já que está inserida na Bacia do rio Paraíba do Sul, uma das bacias hidrográficas federais brasileiras mais relevantes e, ainda, integra o estado do Rio de Janeiro.

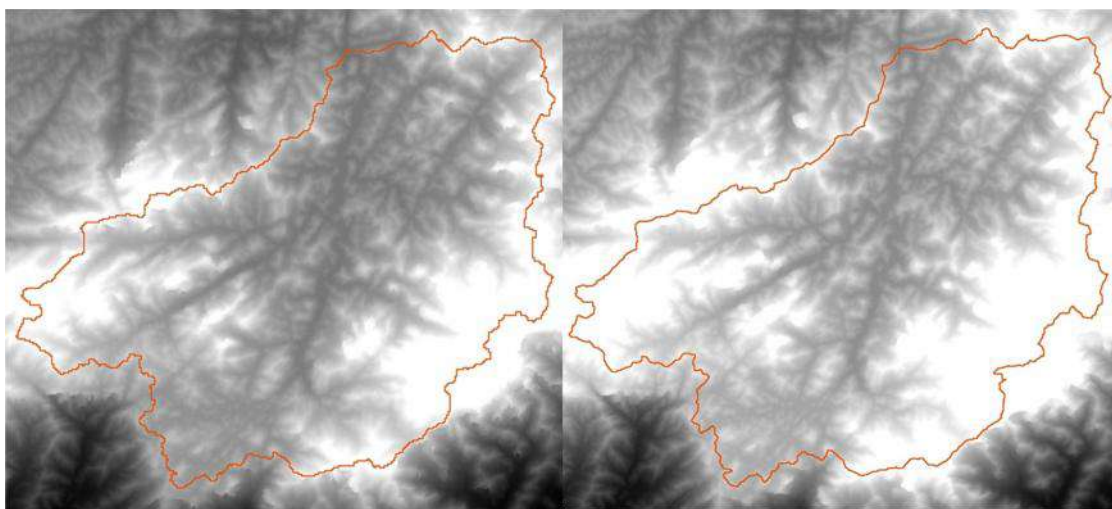
ARAÚJO (2016) fez um levantamento amplo e abrangente da cartografia para a bacia e afirma que é possível fazer uma boa caracterização fisiográfica com base no que está disponível. Entretanto, nota-se que a escala da maioria dos mapas que englobam toda a área da bacia do Piabanha é ainda pequena. O que se observa é o detalhamento de regiões de interesse dentro da bacia como, por exemplo, a base cartográfica da Prefeitura Municipal de Petrópolis na escala de 1:2.000 restrita à região do centro histórico ou o mapa de uso e ocupação do solo gerado a partir da vetorização de ortofotos na escala de 1:25.000 da série de aerolevantamentos de 2005/2006 realizada pelo IBGE em SILVA (2014). Mas, para a bacia como um todo, os levantamentos são pouco detalhados,

Para o estudo desenvolvido foram enfrentados alguns desafios para reunir as informações necessárias conforme será exposto a seguir.

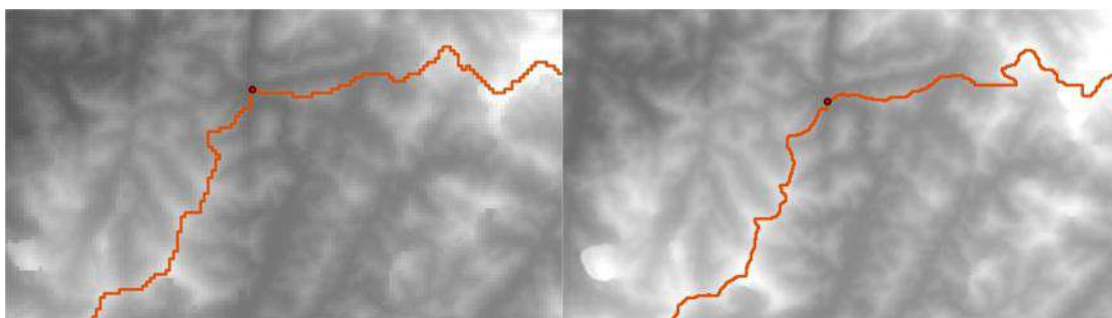
3.4.1 Modelo Digital de Elevação

Modelos Digitais de Elevação (MDE) utilizado foi gerado a partir dos dados SRTM (Shuttle Radar Topography Mission). A SRTM é um projeto internacional liderado pelos Estados Unidos através da NASA e da NGA. Em 2000, a SRTM coletou dados topográficos em quase oitenta por cento da superfície terrestre gerando dados globais de elevação de terra. Num primeiro momento só foram disponibilizados publicamente para outros países os dados com resolução espacial de 90 metros (SRTM-90m). Apenas em 2014 começaram a ser divulgados os dados originais com resolução completa de aproximadamente 30m (SRTM-30m).

Dessa forma, no início das aplicações foi utilizado o SRTM-90m (EMBRAPA, 2013) compatível com a escala 1:25.0000 (IBGE), SRTM-SF-23-Z-B, datum WGS 1984 que foi projetado para UTM, WGS 84, WGS1984 UTM Zone 23S (Figura 24). Com a disponibilização da resolução de 30m as aplicações foram refeitas e os resultados comparados (USGS, 2016). Os dados foram projetados para UTM, WGS 84, WGS1984 UTM Zone 23S e interpolados para preenchimento dos pontos ou áreas sem informação no modelo (Figura 25). As Figuras 26 e 27 representam um zoom da área do exutório da bacia representativa para as duas imagens SRTM (Figuras 24 e 25), com resolução de 30m e 90m respectivamente, onde observa-se a diferença na resolução.



Figuras 24 e 25 – SRTM com resolução espacial de 90m e 30m para a bacia representativa.



Figuras 26 e 27 – Zoom da região do exutório da bacia representativa para demonstrar a diferença de resolução entre o SRTM-90m e SRTM-30m respectivamente.

3.4.2 Mapa de solos

O mapa de solos selecionado foi produzido no âmbito do Projeto Rio de Janeiro que consiste em uma série de estudos multitemáticos do meio físico realizados pelo Serviço Geológico do Brasil – CPRM, na escala 1:250.000, em todo o Estado do Rio de Janeiro em parceria com outras instituições (CARVALHO FILHO *et al.*, 2000). No caso do mapa de solos, o estudo foi desenvolvido em parceria com a Embrapa, DRM-RJ, CIDE, SERLA (atual INEA), EMOP, PUC/RJ, UERJ, RESUB, ANEEL e INMET. O resultado desse trabalho foi um mapa para todo o estado na escala 1:500.000. As unidades de mapeamento foram constituídas por associações, usualmente por duas ou três classes de solos e estão em conformidade com o nível do mapeamento executado (CARVALHO FILHO *et al.*, 2000).

A partir desse projeto, a EMBRAPA (2003) elaborou, em nível de reconhecimento de baixa intensidade um mapa final na escala 1:250.000 com base em fotografias aéreas 1:60.000 (USAF), imagens de satélite Landsat (escala 1:100.000 e 1:250.000) e bases planialtimétricas 1:50.000 (IBGE). Além disso, atualizou a nomenclatura utilizada anteriormente para a classificação dos solos e, ainda, adicionou os percentuais das classes em cada associação na legenda de identificação de solos.

A distribuição espacial dos solos no mapa do estado do Rio de Janeiro é representada em cartas topográficas 1:250.000 através de 161 unidades de mapeamento. Para a bacia representativa, foram observados nove unidades (AR3, AR5, CXbd12, CXbd4, CXbd6, CXbd8, LVAd24, LVAd26, RLd) que combinam, basicamente, as seguintes classes de solo e seus percentuais de distribuição na bacia: Cambissolo

Háplico Tb Distrófico (42%), Latossolo Vermelho-Amarelo (22%), Neossolo Litólico Distrófico (3%) e Afloramento de Rochas (27%) conforme observado na Figura 28.

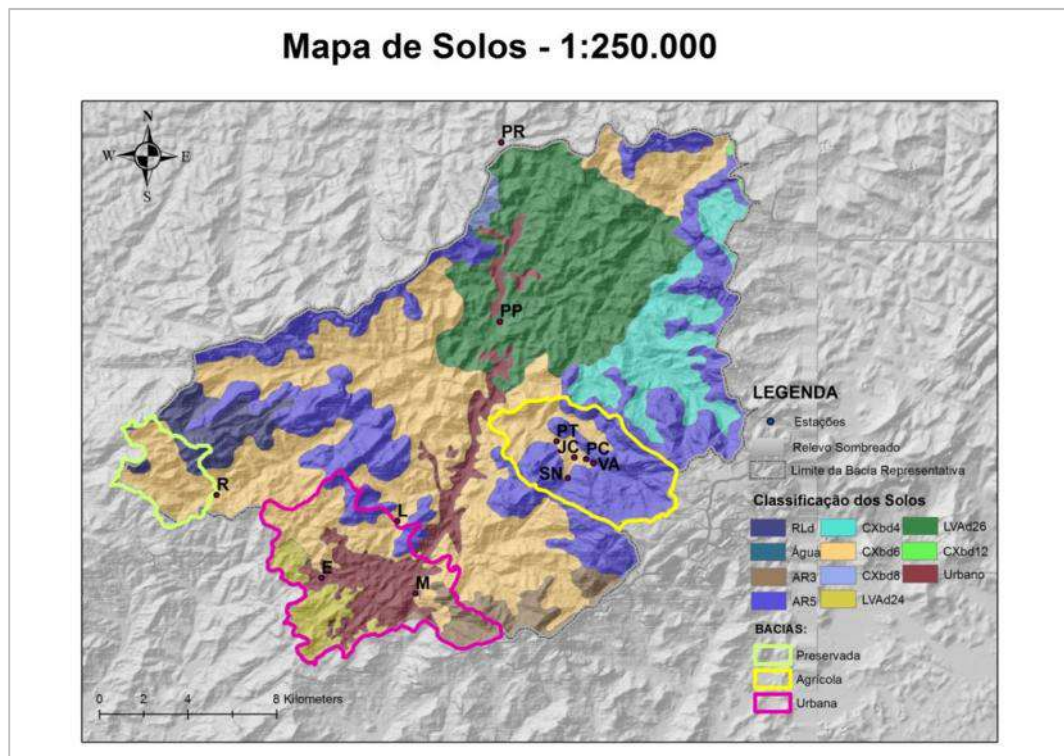


Figura 28 – Mapa de solos para a bacia representativa (adaptado de EMBRAPA, 2003).

3.4.3 Mapa de uso, ocupação de solos e cobertura vegetal

Nesse trabalho foram utilizados dois mapas de uso e ocupação solos em duas diferentes escalas com o objetivo de comparar o impacto de cada um na calibração do modelo. O mapa de menor escala é o elaborado no âmbito do Projeto Rio de Janeiro, na escala 1:250.000, com base na análise digital de imagens de satélite, utilizando-se técnicas de sensoriamento remoto (JACQUES *et* SHINZATO, 2000). Como resultado foram estabelecidas as seguintes classes: Pastagem, Mata, Áreas Urbanas, Solo Exposto, Áreas Agrícolas, Corpos d'Água, Afloramentos de Rocha, Vegetação de Restinga, Campo Inundável, Manguezal, Coberturas Arenosas, Salinas e Extração de Areia. Para a bacia representativa, foram encontradas seis classes conforme apresentado na Figura 29.

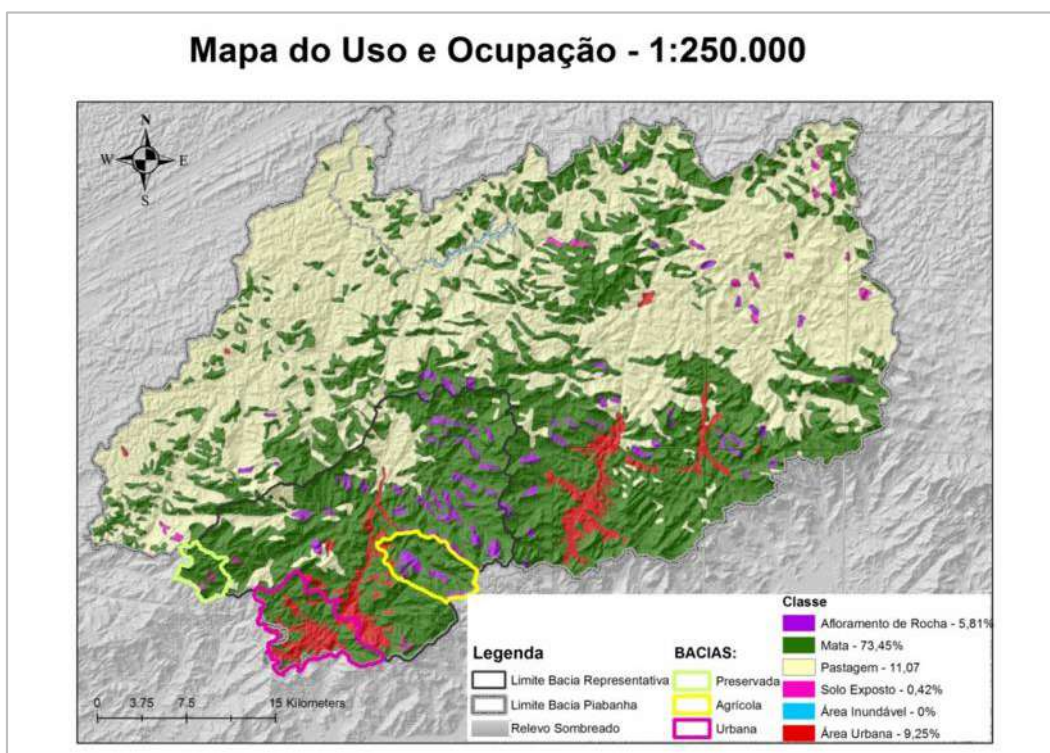


Figura 29. Mapa de Uso e Cobertura do Solo do projeto Rio de Janeiro para a bacia representativa na escala 1:250.000 (adaptado de JACQUES *et* SHINZATO, 2000).

O mapeamento de uso e cobertura para bacia do rio Paraíba do Sul foi elaborado pelo Laboratório ESPAÇO de Sensoriamento Remoto e Estudos Ambientais do Instituto de Geociências (IGeo) da UFRJ em conjunto com o Instituto Estadual do Ambiente (INEA RJ), na escala 1:25.000 com base em Imagens World View ou GeoEye, Imagens RapidEye e Imagens Landsat 5 e 8 e MDE SRTM, com resolução de 30m (INEA, 2017). Nele foram estabelecidas como classes antrópicas: Áreas Antrópicas Não Agrícolas (ANA), Áreas Antrópicas Agrícolas (AAG), identificadas ainda como consolidadas e não consolidadas e Silvicultura (SILV). E como classes naturais: Áreas Naturais Florestadas (ANF), Áreas Naturais Não Florestadas (NNF), Restingas, Manguezais e Corpos D'água. A Figura 30 apresenta o mapa de uso e cobertura para a bacia representativa extraído do mapa da bacia do Paraíba do Sul. É possível observar um melhor detalhamento dos usos gerando um maior número de classes em relação ao mapa do projeto Rio de Janeiro que possui uma escala menor.

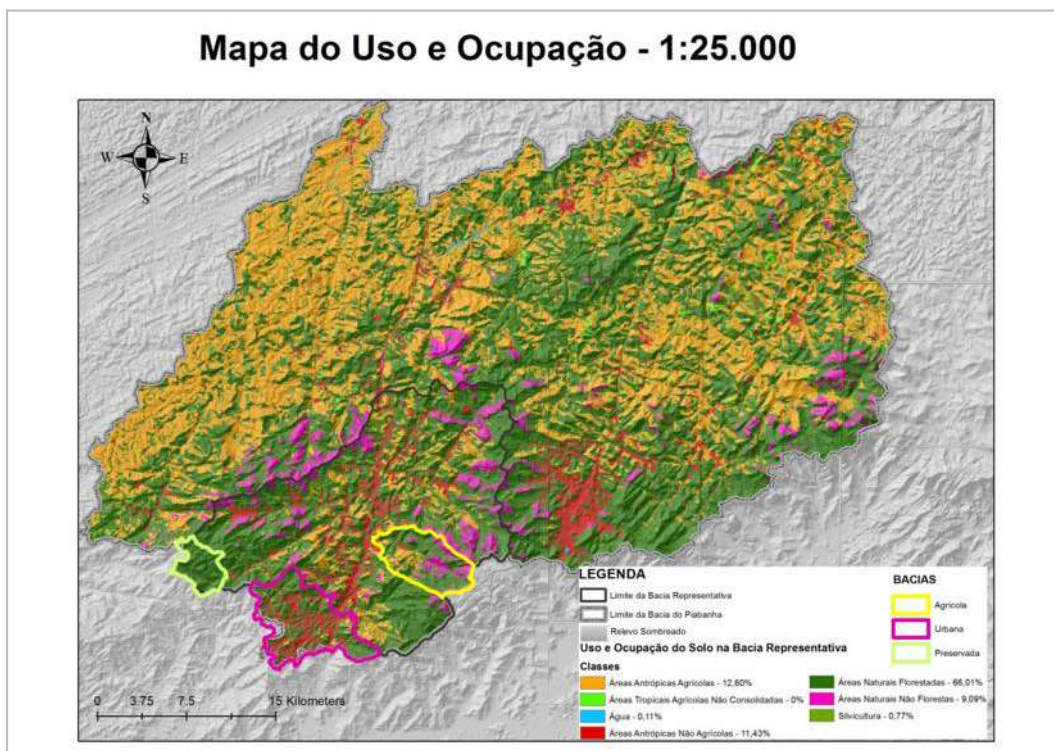


Figura 30. Mapa de uso e cobertura para a bacia representativa, escala 1:25.000. (adaptado de INEA, 2017)

4 Metodologia para avaliação de Redes de Monitoramento de Qualidade da Água com base em Redes Neurais Auto-Associativas

Nesse capítulo primeiramente será apresentada a metodologia proposta para a avaliação da RMQAP. Em seguida a metodologia será aplicada à bacia com base nos dados de qualidade de água disponíveis. Esse capítulo tem como base os resultados publicados em VILLAS-BOAS *et al.* (2017).

4.1 Metodologia

A metodologia proposta tem como base a Análise de Componentes Principais (ACP) e consiste em uma combinação específica de duas técnicas previamente desenvolvidas e já apresentadas na seção 2.5: a Análise de Componente Principal não linear, através das redes neurais autoassociativas - ACPNL/RNAA e a influência geral (IG). Dessa forma, a metodologia será a partir daqui referenciada como ACPNL/RNAA/IG.

Como já discutido no seção 2.5.2, a ACP gera uma transformação ortogonal de um conjunto de variáveis com base em suas correlações lineares cujas novas variáveis, não correlacionadas (i.e. componentes), são combinações lineares das variáveis originais. Elas são plotadas na direção da variabilidade máxima dos dados e ordenadas para que os primeiros componentes, chamados componentes principais, mantenham a maior parte da variação presente nas variáveis originais (JOLLIFFE, 2002). Os *loadings* (i.e. o peso de cada uma das variáveis originais em cada um dos componentes principais) têm sido utilizados, frequentemente, para determinar a importância relativa de uma variável original em um componente, em comparação com outras variáveis, em muitos estudos de qualidade da água. As variáveis de entrada mais relevantes são aquelas com a maior impacto nos componentes principais, ou seja, com os maiores *loadings* (SIMEONOV *et al.*, 2003; SINGH *et al.*, 2004; OUYANG, 2005; SHRESTHA E KAZAMA, 2007; WANG *et al.*, 2012; PHUNG *et al.*, 2015). Isso significa, que as variáveis menos relevantes têm pouco impacto na variabilidade da

qualidade da água e, portanto, podem ser removidas minimizando o impacto na variância dos dados. Dessa forma, a ACP é utilizada para avaliar a relevância de parâmetros e estações de redes de monitoramento de qualidade de água.

Como visto anteriormente, ACPNL/RNAA é uma generalização da ACP tradicional, e tem a capacidade de descrever a estrutura dos dados por subespaços curvos. Assim, os componentes principais podem ser tanto combinações não-lineares quanto lineares das variáveis originais. Os componentes principais dos dados, sem redundâncias e correlações, são extraídos a partir de uma arquitetura de uma rede neural específica conforme visto na seção 2.5.2. Para ACPNL/RNAA, ao contrário do ACP, não é possível indicar diretamente correlações entre variáveis de entrada e componentes principais devido à natureza das redes neurais. Conforme mencionado anteriormente, essas correlações, isto é, os *loadings*, são necessárias para determinar as variáveis mais relevantes de acordo com a ACP. No entanto, a partir da técnica de Influência Geral (IG), apresentada na seção 2.5.1, é possível calcular o nível geral de influência exibido por cada variável de entrada em cada variável de saída de redes neurais a partir de uma análise dos pesos e bias da rede neural utilizada. Assim, traçando um paralelo com a ACP, a IG seria equivalente aos *loadings* das variáveis. Consequentemente, as variáveis com maiores IGs seriam as mais relevantes.

Portanto, a partir da combinação de ACPNL/RNAA e IG é possível determinar a influência das variáveis de entrada em cada componente selecionado. Para isso, propõe-se que valores de IG sejam estimados na camada de estrangulamento, que é responsável por extrair os componentes principais, e não na camada de saída, como a técnica original é tradicionalmente usada com base na Eq. 2. Pois, pretende-se avaliar a relevância das variáveis de entrada (i.e. variáveis originais) em relação aos componentes principais (i.e. novas variáveis). Note-se que a RNAA deve ser "tratada" como uma rede de camada oculta única para calcular a IG, assim, apenas os pesos e bias das camadas de mapeamento e de estrangulamento (ou seja, a primeira e segunda camadas ocultas) são utilizados para o cálculo da IG. Na Figura 8, que apresenta um exemplo de arquitetura de RNAA com entradas (I_n), saídas (O_n), neurônios representados por círculos, camadas ocultas de neurônios, e pesos e bias (w_{ji} , v_{li}) representados por setas. Os pesos e bias usados para calcular as IGs são as setas em negrito. A camada de estrangulamento (círculo preto) possui um neurônio apenas, ou seja, um componente principal. Dessa

forma, nesse exemplo, a aplicação da metodologia ACPNL/RNAA/IG resultaria na avaliação da relevância das cinco variáveis de entrada (I_1 - I_5) no único componente principal. Ao final teríamos o percentual de influência de cada uma das variáveis no componente principal.

Quando a relevância dos parâmetros da qualidade da água deve ser avaliada, a entrada ACPNL/RNAA/IG consiste em uma matriz de concentrações observadas em que as linhas correspondem a todas as combinações diferentes de locais de amostragem e datas de amostragem e as colunas às diferentes águas parâmetros de qualidade (Tabela 4); e os resultados são as IGs que representam a relevância de cada parâmetro de qualidade da água. Quando a relevância dos locais de amostragem deve ser avaliada, a entrada ACPNL/RNAA/IG consiste em uma matriz de concentrações observadas em que as linhas correspondem a todas as diferentes combinações de parâmetros de qualidade da água e datas de amostragem, e as colunas às diferentes amostras Localizações; e os resultados são as IGs que representam a relevância de cada local de amostragem.

4.2 Estudo de Caso

A metodologia proposta foi aplicada à bacia representativa do rio Piabanha como estudo de caso.

4.2.1 Seleção dos dados

A Rede de Qualidade de Água da Bacia Representativa do rio Piabanha (RMQAP) começou a ser operada em agosto de 2009 através da realização de campanhas de coletas de amostras e medição com equipamento automático mensalmente, a partir de 2010, a frequência passou a ser bimestral e em 2013 voltou a ser mensal. A RMQAP é formada por nove estações de monitoramento distribuídas ao longo da bacia representativa sendo três delas na bacia urbana, três na bacia rural, uma na bacia de mata preservada e duas estações fora das bacias experimentais, no curso principal, chamadas de estações de controle. A Figura 31 apresenta a bacia representativa, as bacias experimentais, as estações de qualidade de água e alguns dos

usuários instalados na bacia com base em dados fornecidos pelo INEA – Instituto Estadual do Ambiente e pela ANA – Agência Nacional de Águas através do CNARH – Cadastro Nacional de Usuários de Recursos Hídricos. Com base na relação de usuários de lançamento, foi possível constatar que a maior parte deles utiliza fossa e sumidouros. Já as poucas indústrias que lançam diretamente nos corpos d'água, em sua maioria fazem algum tipo de tratamento no efluente antes de lançá-lo. (VILLAS-BOAS *et al.*, 2011).

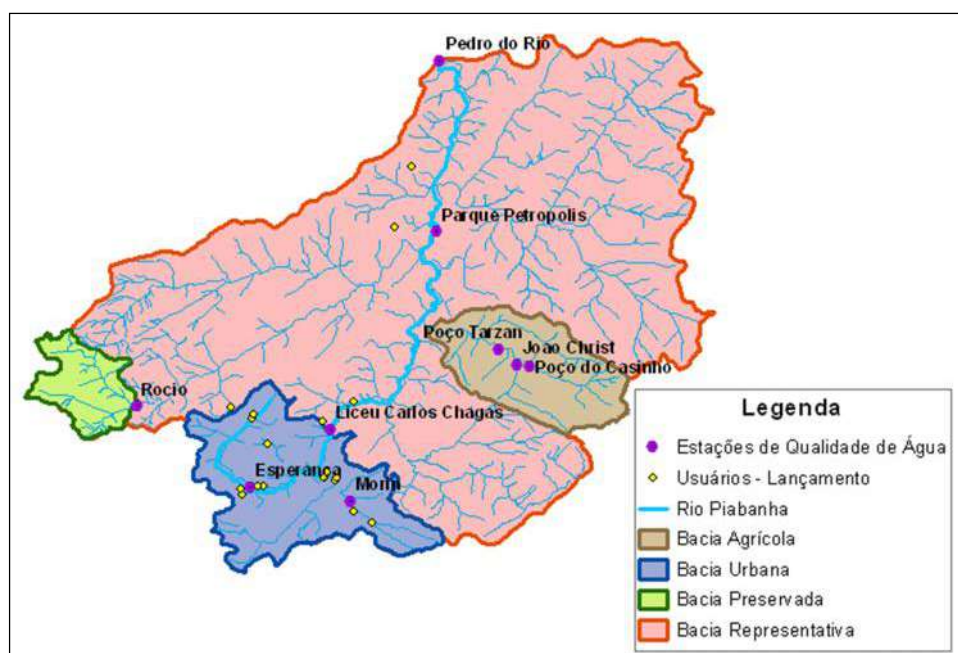


Figura 31 - Localização das estações de qualidade de água selecionadas para o estudo e dos usuários de lançamento na bacia Representativa (VILLAS-BOAS *et al.*, 2011).

Vale ressaltar que o objetivo da RMQAP é avaliar o impacto dos diferentes usos e ocupação do solo na bacia representativa e que ela foi instalada sem seguir nenhum protocolo de planejamento de rede. Assim, as informações geradas nas bacias experimentais urbana e agrícola podem ser comparadas com as da bacia preservada (*background*) e o impacto pode ser avaliado. O critério adotado para a definição dos pontos de amostragem foi a existência prévia de estação fluviométrica de forma a garantir que houvesse dados de quantidade (cotas e vazão) associados aos dados de qualidade da água. Os parâmetros analisados foram selecionados com base nos mapeamentos de uso e ocupação do solo e na disponibilidade do laboratório da CPRM. A frequência de coleta foi definida de acordo com o orçamento do projeto que sofre

constantes cortes e contingenciamentos. A Tabela 2 apresenta a relação das estações RMQAP com suas respectivas características.

Tabela 2 - Relação das estações de qualidade de água da RMQAP.

ESTAÇÃO	CÓDIGO	CURSO D'AGUA	LATITUDE	LONGITUDE	Bacia Exp.
Rocio 2 (R)	58400212	Rio da Cidade	22° 28' 39"	43° 15' 25"	Preservada
Esperança (E)	58400010	Rio Piabanha	22° 30' 39"	43° 12' 37"	Urbana
Liceu (L)	58400050	Rio Piabanha	22° 29' 14"	43° 10' 38"	Urbana
Morin (M)	58400030	Rio Palatinado	22° 31' 00"	43° 10' 08"	Urbana
Poço Tarzan (PT)	58400110	Rio Bonfim	22° 27' 14"	43° 06' 28"	Agrícola
Poço Casinho (PC)	58400104	Rio Açu	22° 27' 40"	43° 05' 41"	Agrícola
Joao Christ (JC)	58400108	Rio Alcobaça	22° 27' 38"	43° 06' 00"	Agrícola
Pq.Petropolis (PP)	58400250	Rio Quitandinha	22° 24' 19"	43° 08' 00"	-
Pedro do Rio (PR)	58405000	Rio Piabanha	22° 19' 56"	43° 08' 01"	-

São analisados cerca de 30 parâmetros de qualidade de água pelo Laboratório de Análise de Águas Minerais – LAMIN, da CPRM, e Laboratório contratado (para os parâmetros que o LAMIN não realiza a análise) apresentados na Tabela 3. Com equipamento automático é feita a medição de Temperatura, pH, Condutividade Elétrica e Oxigênio Dissolvido. Os dados selecionados para o estudo são relativos as cammanhas realizadas em agosto, setembro, outubro, novembro e dezembro de 2009; Março, junho, julho e novembro de 2010; Fevereiro, junho, agosto e dezembro de 2011; e março, maio e agosto de 2012. As amostras foram coletadas em diferentes momentos do dia e/ou diferentes dias do mês para cada parâmetro.

Tabela 3 - Lista de Parâmetros da rede de qualidade de água da área de estudo.

Parâmetros - LAMIN			Laboratório Privado	Sonda Multiparamétrica
Alumínio (Al)	Condutividade Elétrica - 20°C(CE)	pH	DBO	pH
Antimônio (Sb)	Estanho (Sn)	Prata (Ag)	DQO	Temperatura (T)
Arsênio (As)	Ferro (Fe)	Selênio (Se)	Organoclorados (SVOC)	Condutividade Elétrica - 20°C(CE)
Cádmio (Cd)	Fosfato Total (PO_4^{-3})	Sólidos em Suspensão	Organofosforados	Oxigênio Dissolvido (OD)
Chumbo (Pb)	Manganês (Mn)	Sulfato (SO_4^{-2})	Carbamatos	
Cobalto (Co)	Níquel (Ni)	Turbidez (Tb)	Coliformes Fecais (CF)	
Cobre (Cu)	Nitrato (NO_3^-)	Zinco (Zn)	Coliformes Totais (CT)	
Cromo Total (Cr)	Nitrogênio Amoniacal Total (NH4)			

4.2.2 Pré- Processamento dos dados

Os dados de qualidade da água utilizados neste estudo são dados reais e, por isso, são altamente suscetíveis a ruído, falhas e inconsistências que podem distorcer os resultados (HAN *et al.*, 2011) e, consequentemente, devem ser analisados preliminarmente antes da aplicação da metodologia.

O banco de dados disponível para este estudo equivale a valores medidos de 30 parâmetros de qualidade da água em 9 locais de amostragem e 16 datas diferentes. Os dados foram organizados em forma de uma matriz, onde as colunas correspondem às *variáveis* e, as linhas, às *observações* ou *registros*, necessária para a aplicação da metodologia proposta. As variáveis compreendem as estações e parâmetros de monitoramento. A Tabela 4 exemplifica a matriz de entrada de dados tendo como variável os parâmetros de qualidade de água.

Foi realizada a caracterização dos dados através dos histogramas, box-plots e coeficientes de assimetria, bem como o cálculo dos coeficientes de correlação das variáveis: *r* de Pearson e *Rho* de Spearman. *R* de Pearson é um coeficiente amplamente utilizado em abordagens estatísticas. Ele representa uma medida de associação linear entre variáveis e assume que os dados seguem uma distribuição normal. Por outro lado, *Rho* de Spearman é uma medida estatística não paramétrica de dependência, de modo que não é influenciada pela distribuição subjacente. Em vez disso, ele mede a força de dependência linear entre os registros de dados e é frequentemente usado em estudos de qualidade da água (MC BRIDE *et al.*, 2005).

Embora algumas estações possuam períodos de registro mais longos, o conjunto de dados foi limitado ao período 2010-2012 para evitar distorções geradas por mudanças nas condições hidrológicas e/ou ambientais. Os registros com falhas foram eliminados (HAN *et al.*, 2011a; TANOS *et al.*, 2015). Na Tabela 4, por exemplo, as observações número 1 e n têm falhas e precisam ser eliminadas. Em todo o conjunto de dados, apenas a temperatura e o oxigênio dissolvido apresentavam falhas. Vale ressaltar que os dados eram brutos e não passaram por nenhum procedimento de consistência específico.

Tabela 4 – Exemplo de matriz de entrada de dados para metodologia proposta.

Obs. #	Estação de Monitoramento (Parâmetro de Qualidade da Água) - Data	Variável					
		Parâmetro de Qualidade da Água (Estação de Monitoramento)					
		Temperatura	OD	Al	Coliforme Fecal	...	m
1	Liceu - Mar/2010			0,003	16000	...	7,70
2	Liceu - Jun/2010	18,58	4,86	0,003	920	...	4,43
3	Liceu - Out/2010	18,89	5,03	0,003	45410	...	5,43
...
n	Rocio - Ago/2012			0,015	131	...	4,90

* Observações com falhas em negrito.

Para esta primeira análise, os valores abaixo do limite de quantificação (QL) do método de análise, denominados não-detectados (APHA, 2005), foram substituídos pelo valor QL. De acordo com a EPA (2009), o procedimento para analisar dados não-detectados depende da porcentagem de valores não detectados. Para métodos de substituição, a EPA (2009) recomenda 15% ou menos de não- detectados. Cerca de metade dos parâmetros tinham mais de 80% dos valores não-detectados e foram excluídos da análise. Portanto, 13 parâmetros com menos de 15% de não-detectados foram utilizados na análise (parâmetros em negrito na Tabela 3. Como referência, OISEN *et al.* (2012) testaram a substituição desses valores por zero, metade e uma vez a QL em um conjunto de dados com menos de 10% dos não-detectados e encontrou resultados semelhantes nos três casos. O software LOADFLEX, desenvolvido pelo Serviço Geológico americano (USGS) substitui os valores não-detectados pelo valor do limite de detecção (RUNKEL *et al.*, 2004). A Tabela 5 apresenta os valores máximos,

mínimos, média e mediana para os parâmetros de qualidade de água selecionados para as bacias experimentais e a representativa.

Tabela 5- Valores máximos, mínimos, média e mediana para os parâmetros de qualidade de água selecionados para as bacias experimentais e a representativa.

Parâmetros de Qualidade de Água	Bacia urbana				Agrícola				Preservada				Representativa			
	Média	Mediana	Max	Min	Média	Mediana	Max	Min	Média	Mediana	Max	Min	Média	Mediana	Max	Min
Al	0.014	0.019	0.094	0.003	0.063	0.076	0.300	0.014	0.024	0.027	0.072	0.006	0.024	0.044	0.300	0.003
NH4+	1.600	3.910	24.000	0.025	0.025	0.049	0.200	0.025	0.025	0.045	0.100	0.025	0.050	1.382	24.000	0.025
DQO	30	34	79	7	8	9	15	5	9	10	26	5	15	19	79	5
OD	3.44	3.80	6.86	0.40	7.70	7.19	8.94	4.96	7.84	6.74	8.58	2.20	5.73	5.82	9.39	0.40
CE	163.7	253.0	737.0	22.9	25.1	30.0	223.0	8.8	15.8	15.9	18.6	13.5	91.8	118.5	737.0	8.8
CF	31950	111131	816400	16	305	405	1414	17	1641	92774	866400	41	1995	50159	866400	16
Fe	0.043	0.065	0.257	0.010	0.029	0.038	0.154	0.009	0.025	0.026	0.041	0.005	0.035	0.050	0.257	0.005
NO3-	2.80	5.63	22.41	0.01	2.94	2.76	9.76	0.01	0.59	0.58	1.02	0.15	3.20	5.03	22.41	0.01
pH	6.59	6.48	7.03	5.53	6.73	6.63	7.09	3.35	6.78	6.79	6.97	6.55	6.64	6.55	7.09	3.35
SO42-	8.69	20.03	116.92	4.94	0.73	1.77	32.04	0.26	0.56	0.60	0.76	0.47	4.54	8.51	116.92	0.26
Temp	20.39	21.05	26.99	16.82	18.10	18.00	22.30	13.73	17.51	17.11	19.58	14.12	19.36	19.53	27.32	13.73
CT	98590	1287217	17329000	540	4981	13129	86640	240	14048	187339	1095000	540	21767	498508	17329000	240
Turb	5.72	7.74	36.00	0.02	2.21	2.75	10.70	0.20	0.59	0.82	1.60	0.20	4.32	6.88	85.30	0.02

Os *outliers* podem ser gerados por um mau funcionamento do equipamento ou por um evento hidrológico raro. Enquanto, no primeiro caso, os valores anuais devem ser eliminados, no último, eles são de fundamental importância hidrológica. No entanto, enquanto os *outliers* caem dentro de valores razoáveis, é difícil saber de qual tipo eles são. Nesse caso específico, não havia nenhum sinal de problema de equipamento e os valores, embora extremos, estavam dentro de limites razoáveis. Foi feita uma análise desses valores com base num método não-paramétricos chamado matriz de distâncias. Consequentemente, embora para a análise seja recomendável a eliminação de *outliers*, as medições contendo valores um pouco altos não foram removidas do conjunto de dados para esta primeira aplicação.

Para comparar os valores de todos os parâmetros, eles foram normalizados usando a transformação Z-score para o PCA e a transformação Min-Max para o intervalo [-1, 1] para a NLPCA.

4.2.3 Aplicação da metodologia

A metodologia foi aplicada em duas partes: primeiro, considerando os parâmetros da qualidade da água como as variáveis cuja relevância deve ser avaliada e, seguir, considerando as estações de monitoramento. Para cada parte, primeiro, a relevância foi avaliada usando o ACP tradicional e, em seguida, a metodologia proposta, ACPNL/RNAA/IG. O desempenho das duas técnicas foi comparado a partir de métricas estatísticas. O software MATLAB R2012a. foi utilizado.

Para ACP, foi aplicada rotação Varimax aos eixos dos componentes principais distribuindo assim os valores dos *loadings* de modo a torna-los próximos de zero ou um. É uma forma de simplificar a estrutura dos fatores e facilitar a interpretação desses a partir da redução de algumas ambiguidades que geralmente acompanham a solução inicial não rotacionada (HAIR et al., 2009; PALLANT, 2010). Assim, idealmente a variância de cada parâmetro passa a ser explicada, idealmente, por um único componente principal, o que simplifica a interpretação dos resultados (SELLE et al., 2013; SIMEONOVA et al., 2003; OUYANG et al., 2006). Foram considerados três critérios para selecionar o número de componentes principais relevantes: (1) Porcentagem de Variação Total (JOLLIFFE, 2002), (2) *Kaiser's rule* (KAISER, 1960) e (3) *Scree Test* (CATTELL, 1966). No primeiro, a variância dos componentes relevantes

tem de ser maior do que 70% da variância da amostra (OUYANG, 2005; JOLLIFFE, 2002). Na *Kaiser's rule*, os componentes relevantes são aqueles que têm autovalores superiores a um (KAISER, 1960). Este critério tem sido amplamente aplicado em estudos de qualidade da água (MAGYAR *et al.*, 2013; PAGE *et al.*, 2012; SELLE *et al.*, 2013). No *Scree test*, os números dos componentes são plotados contra os seus valores correspondentes e os componentes relevantes são definidos no ponto de inflexão na curva (VEGA *et al.*, 1998; MISHRA, 2010). Este último critério tem uma parcela de análise um tanto subjetiva, portanto, foi utilizado apenas de forma acessória aos demais. As variáveis com *loadings* superiores a 0,70 foram consideradas relevantes (SIMEONOV *et al.*, 2003; SINGH *et al.*, 2004; OUYANG, 2005; SHRESTHA *et al.*, 2007) para garantir que pelo menos metade da variação variável seja explicada pelo componente (HAIR *et al.*, 2009).

Para ACPNL/RNAA/IG, a arquitetura RNAA incluiu uma rede neural *feedforward* com a função tangente-sigmoide nas camadas ocultas e uma função linear na camada de saída. O treinamento da RNAA usou o algoritmo *backpropagation*, e os valores iniciais de pesos e bias iniciais foram determinados usando o método *Nguyen-Widrow* (NGUYEN E WIDROW, 1990). A regularização bayesiana (MACKAY, 1992) foi usada para evitar *overfitting*, o que também melhorou a generalização da rede neural.

A seleção da melhor arquitetura para a AANNs foi feita a partir do teste de diferentes números de neurônios para as camadas ocultas com base no método chamado “Sub-amostragem aleatória repetida” ou “Validação de Sub-amostragem aleatória” (HAN *et al.*, 2011), no qual o conjunto de dados foi aleatoriamente e repetidamente subdividido em um conjunto de dados de treinamento, contendo 70% de os dados originais, e um conjunto de dados de validação, que contém os 30% restantes. Para cada divisão, a rede foi ajustada para o conjunto de dados de treinamento, e sua acurácia preditiva foi avaliada utilizando o conjunto de dados de validação. A acurácia preditiva de cada arquitetura de rede foi determinada utilizando a média do EMQ, do Erro de Predição Final (EPF) e do Critério Teórico de Informação (CTI) calculados para cada uma de suas divisões. Para cada divisão, o EPF é dado pela Eq. 8 (KRAMER, 1991):

$$EPF = e^{\frac{1 + \frac{d}{m n}}{1 - \frac{d}{m n}}} \quad (8)$$

e o CTI pela Eq. 9 (KRAMER, 1991):

$$CTI = \ln(e) + 2 \frac{d}{m n} \quad (9)$$

onde $e = EMQ/2$, $d = (m + f + 1)(M1 + M2) + m + f$ que é número de pesos, f é o número de neurônios na camada de estrangulamento, $M1$ é o número de neurônios na camada de mapeamento e $M2$ é o número de neurônios na camada de desmapeamento the number of neurons in the demapping layer.

Estes valores expressam a relação de troca entre a precisão do ajuste e número de variáveis ajustáveis em termos explícitos (KRAMER, 1991). Note-se que o número de neurônios em cada arquitetura testada tinha, ainda, que satisfazer condição de KRAMER (1991) estabelecida na Eq. 10:

$$M1 + M2 \ll m(n - f)/(m + f + 1) \quad (10)$$

Além disso, e também de acordo com KRAMER (1991), foi definido o mesmo número de neurônios nas camadas de mapeamento e desmapeamento ($M1 = M2$) por simplificação. Após a seleção da arquitetura foram definidos os pesos e bias finais para todo o conjunto de dados. O treinamento foi realizado várias vezes na tentativa de encontrar mínimo global da função objetivo (KRAMER, 1991). Com os pesos e bias finais foi calculada a IG de cada variável.

Por último, os desempenhos das duas técnicas, ACP e ACPNL/RNAA/IG foram comparados a partir dos valores da Fração Explicada da Variância (FEV) e da capacidade de reconstrução dos dados originais de cada uma usando EPF, CTI, EMQ e o coeficiente de determinação (R^2) (KRAMMER, 1991; MONAHAN, 2000; NEWBIGGING, 2003) given by:

$$FEV = 1 - \left| \frac{\sum_{i=1}^m Var(t_i) - \sum_{i=1}^m Var(a_i)}{\sum_{i=1}^m Var(t_i)} \right| \quad [11]$$

onde o numerador é a variância dos dados originais menos a variância dos dados transformados, ou seja, após a aplicação da análise de componente principal, dividido pela variância dos dados originais.

4.2.4 Resultados

Parâmetros de Qualidade de Água

O conjunto de dados incluiu os valores transformados de 13 parâmetros de qualidade da água (PQA) como variáveis e 70 observações para cada um desses parâmetros. A Figura 32 apresenta os box-plots para os PQA. É possível verificar diversos pontos extremos (outliers – cruz vermelha) e que a maioria das variáveis apresentaram distribuições assimétricas com o deslocamento da mediana (linha vermelha) e também a partir dos valores do coeficiente de assimetria apresentados na Figura 33 através do gráfico de barras. Além disso, observou-se que a maioria dos parâmetros apresentava distribuições assimétricas (Tabela 3) e precisava passar adicionalmente por uma transformação adicional usando logarítmo para reduzir a assimetria, (OLSEN *et al.*, 2012).

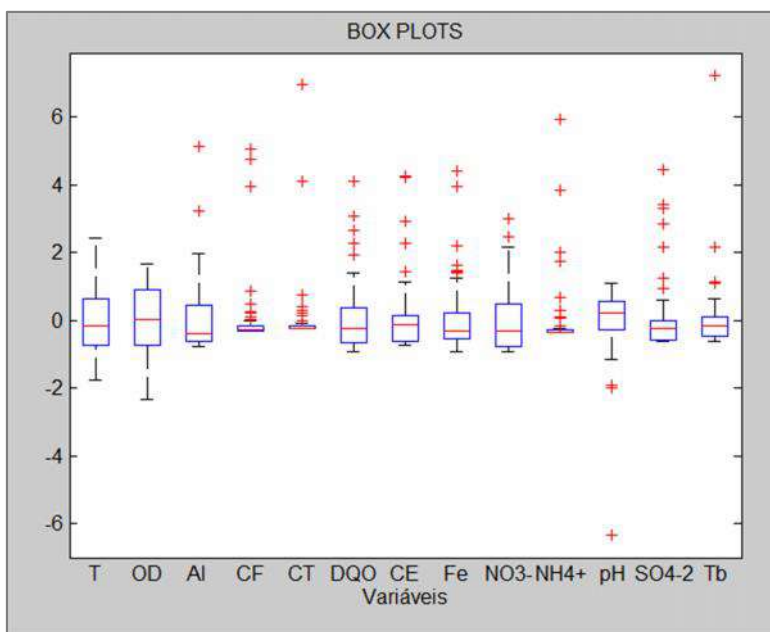


Figura 32 – Box-plot para os parâmetros de qualidade de água.

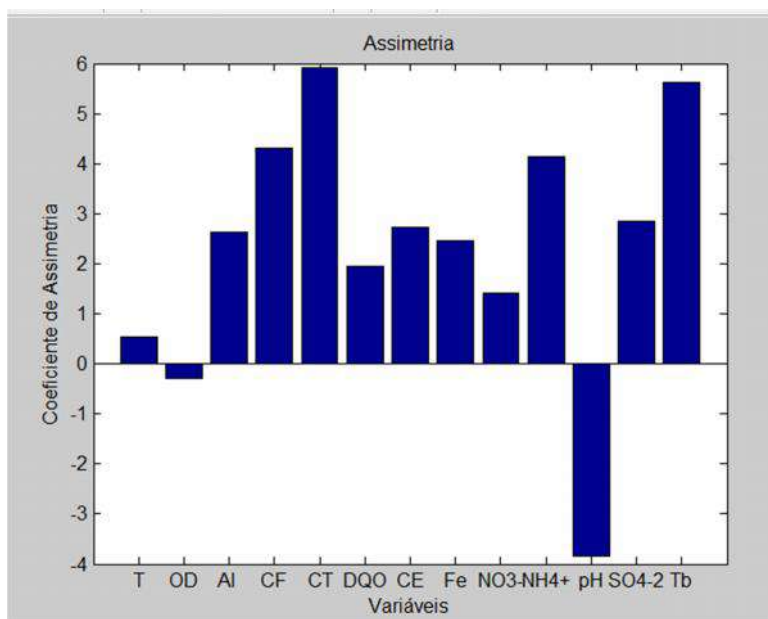
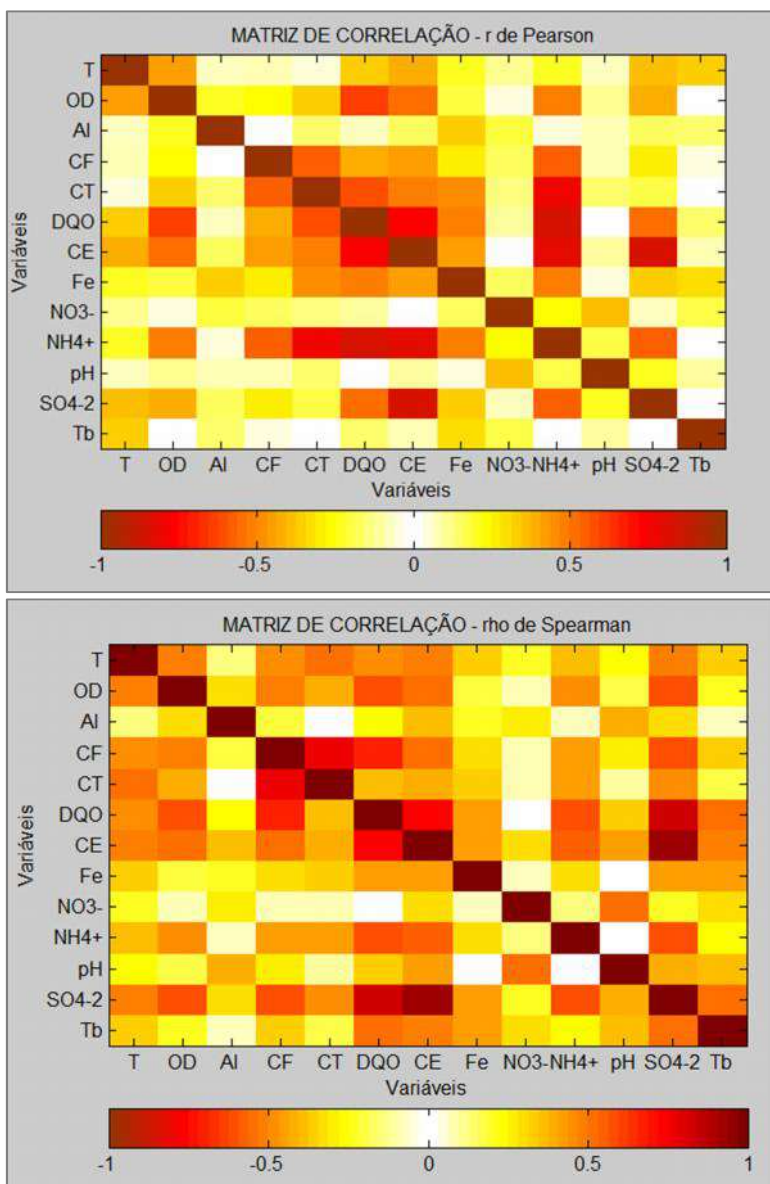


Figura 33 - Coeficiente de Assimetria para os parâmetros de qualidade da água.

As Figuras 34 e 35 apresentam os resultados para as matrizes de correlação para os parâmetros de qualidade de água com base nos coeficientes r de Pearson e ρ de Spearman respectivamente. Os resultados mostraram uma correlação fraca (ou seja, $|r| < 0.3$ ou $|\rho| < 0.3$) para 60% dos pares para r e 38 % para ρ ; e correlação moderada (ou seja, $0.3 < |r| < 0.5$ ou $0.3 < |\rho| < 0.5$) para 23% dos pares para r e 38% para ρ . Os intervalos de r e ρ foram definidas em Cohen (1988). O teste de Bartlett's teve significância aproximadamente zero para todas as variáveis.. É possível notar que os parâmetros Amônia (NH_4^+), Coliformes Fecais e Totais, DQO e Condutividade Elétrica apresentam boa correlação entre si. Já o parâmetro Turbidez possui apresentou baixa correlação com quase todos os parâmetros.

O *ACP* resultou em quatro componentes principais. Os quatro primeiros componentes explicaram 76% da variância contida no conjunto de dados original e foram os únicos com autovalores superiores a um (Tabela 4). A Tabela 5 mostra os *loadings* de cada parâmetro de qualidade da água após a aplicação da rotação *Varimax* para os componentes selecionados. Os parâmetros relevantes baseados no *ACP* são: Alumínio (*Al*), Amônia (NH_4^+), Demanda Química de Oxigênio (DQO), Oxigênio Dissolvido (OD), Coliformes Fecais (CF), Coliformes Totais (CT) e Turbidez (Turb) (valores de *loadings* apresentados em negrito na Tabela 5).



Figuras 34 e 35 – Matriz de correlação para os parâmetros de qualidade de água com base nos coeficientes r de Pearson e ρ de Spearman respectivamente.

Ao calcular o r de *Pearson* e o ρ de *Spearman* para cada um dos 78 pares de parâmetros possíveis (ou seja, combinações de 13 parâmetros tomados 2 de cada vez), os resultados mostraram uma correlação fraca (ou seja, $|r| < 0.3$ or $|\rho| < 0.3$) para 60% dos pares para r e 38 % para ρ ; e correlação moderada (ou seja, $0.3 < |r| < 0.5$ or $0.3 < |\rho| < 0.5$) para 23% dos pares para r e 38% para ρ . Os intervalos de r e ρ foram definidas em Cohen (1988).

O *PCA* resultou em quatro componentes principais. Os quatro primeiros componentes explicaram 76% da variância contida no conjunto de dados original e

foram os únicos com autovalores superiores a um (Tabela 6). A Tabela 7 mostra os pesos de cada parâmetro de qualidade da água após a aplicação da rotação *Varimax* para os componentes selecionados. Os parâmetros relevantes baseados no *PCA* são: Alumínio (*Al*), Amônia (*NH₄⁺*), Demanda Bioquímica de Oxigênio (COD), Oxigênio Dissolvido (DO), Coliformes Fecais (FC), Coliformes Totais (TC) e Turbidez (Turb) (valores dos *loadings* apresentados em negrito na Tabela 7).

Tabela 6 - Os autovalores e os percentuais de variância definidos para cada componente principal.

Componentes	Auto Valores	Variância (%)
1	5.6	42.8
2	1.9	14.5
3	1.4	10.4
4	1.1	8.1
5	0.8	6.5
6	0.5	4.0
7	0.5	3.8
8	0.4	3.3
9	0.3	2.6
10	0.2	1.7
11	0.1	1.1
12	0.1	0.7
13	0.0	0.3

Tabela 7 – Loadings das 13 variáveis de qualidade de água para os quatro componentes principais após a rotação Varimax aplicada aos parâmetros de qualidade de água.

Qualidade de Água parâmetros/componentes	Loadings			
	1	2	3	4
Al	-0.13	-0.88	0.03	-0.07
NH4+	0.78	-0.03	0.18	0.36
DQO	0.72	0.14	0.45	0.32
OD	-0.74	-0.16	-0.01	-0.40
CE	0.44	0.42	0.61	0.38
CF	0.42	0.13	0.10	0.75
Fe	0.25	-0.43	0.62	0.29
NO3-	-0.64	0.49	0.25	0.32
pH	0.06	-0.40	-0.53	0.15
SO42-	0.49	0.40	0.62	0.37
Temp	0.07	0.11	0.32	0.69
CT	0.29	-0.11	0.01	0.86
Turb	0.02	-0.08	0.82	0.14

Na ACPNL/RNAA/IG, foram consideradas diferentes arquiteturas de rede com valores crescentes de f a partir de um, e valores de $M1$ e $M2$ começando em dois e aumentando a cada dois. A Tabela 8 mostra os valores de MSE, FPE e AIC para as diferentes combinações de f e $M1$ e $M2$. Pode-se observar que o melhor ajuste, definidos por valores mais baixos de EMQ, EPF e CTI, correspondem a $f = 3$ e $M1 = M2 = 6$. O IG de cada variável é mostrado na Tabela 9.

Tabela 8 - Principais arquiteturas de redes neurais para NLPCA para avaliação de parâmetros de qualidade da água.

	f=1			f=2			f=3			f=4		
M1/M2	EMQ	EPF	CTI	EMQ	EPF	CTI	EMQ	EPF	CTI	EMQ	EPF	CTI
8	0.117	0.016	-4.146	0.137	0.020	-4.119	0.073	0.011	-4.689	0.058	0.009	-4.951
6	0.096	0.012	-4.533	0.079	0.010	-4.673	0.048	0.006	-5.251	0.068	0.009	-5.051
4	0.111	0.012	-4.495	0.102	0.011	-4.557	0.074	0.008	-4.922	0.091	0.010	-4.663

Tabela 9 - O IG (%) calculado para os parâmetros de qualidade de água.

Parâmetros de Qualidade de Água	GI (%)
Coliformes Fecais (CF)	13
Nitrato (NO ₃ -)	11
Temperatura (Temp)	10
Oxigênio Dissolvido (OD)	9
Sulfato (SO ₄ ²⁻)	9
Ferro (Fe)	9
Amônia (NH ₄ ⁺)	7
Alumínio (Al)	6
pH	6
Condutividade Elétrica (CE)	6
Coliformes Totais (CT)	6
Turbidez (Turb)	4
Demanda Química de Oxigênio (DQO)	4

Estações de Monitoramento

O conjunto de dados foi organizado considerando os dados transformados das nove estações de monitoramento como variáveis de entrada com 122 observações disponível para cada uma dessas variáveis. A matriz de entrada de dados para os pontos de amostragem ficou com 9 colunas (variáveis) e 122 linhas (observações). Aqui também foi possível verificar a presença de *outliers* e as distribuições assimétricas das variáveis com base nos box-plots e coeficiente de assimetria. A Figura 36 apresenta os box-plots para as estações de monitoramento de qualidade de água (EMQA) e a Figura 37 o coeficiente de assimetria.

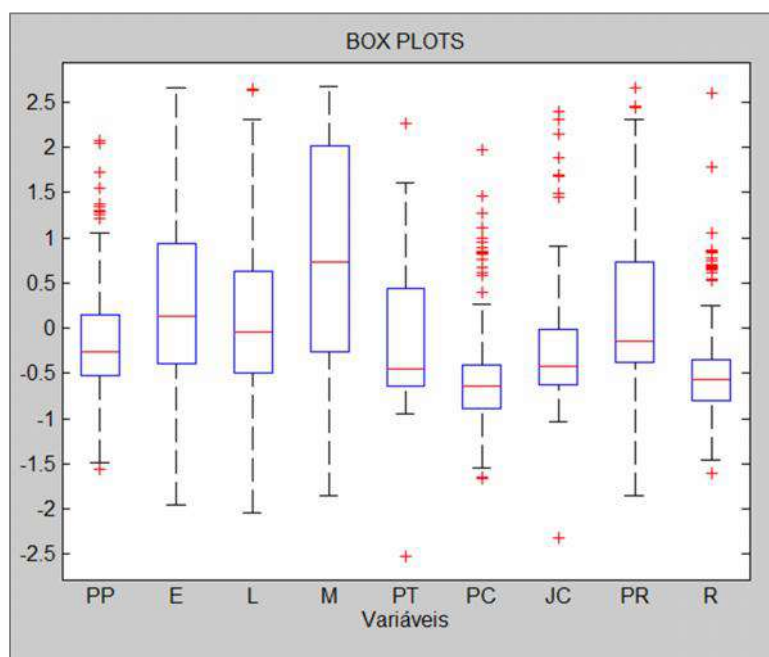


Figura 36 – Box-plots para as estações de monitoramento de qualidade de água.

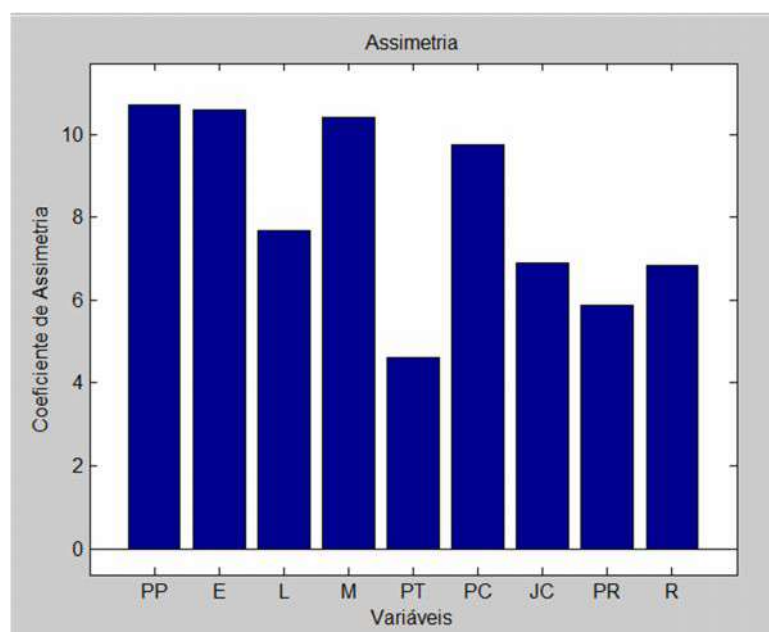
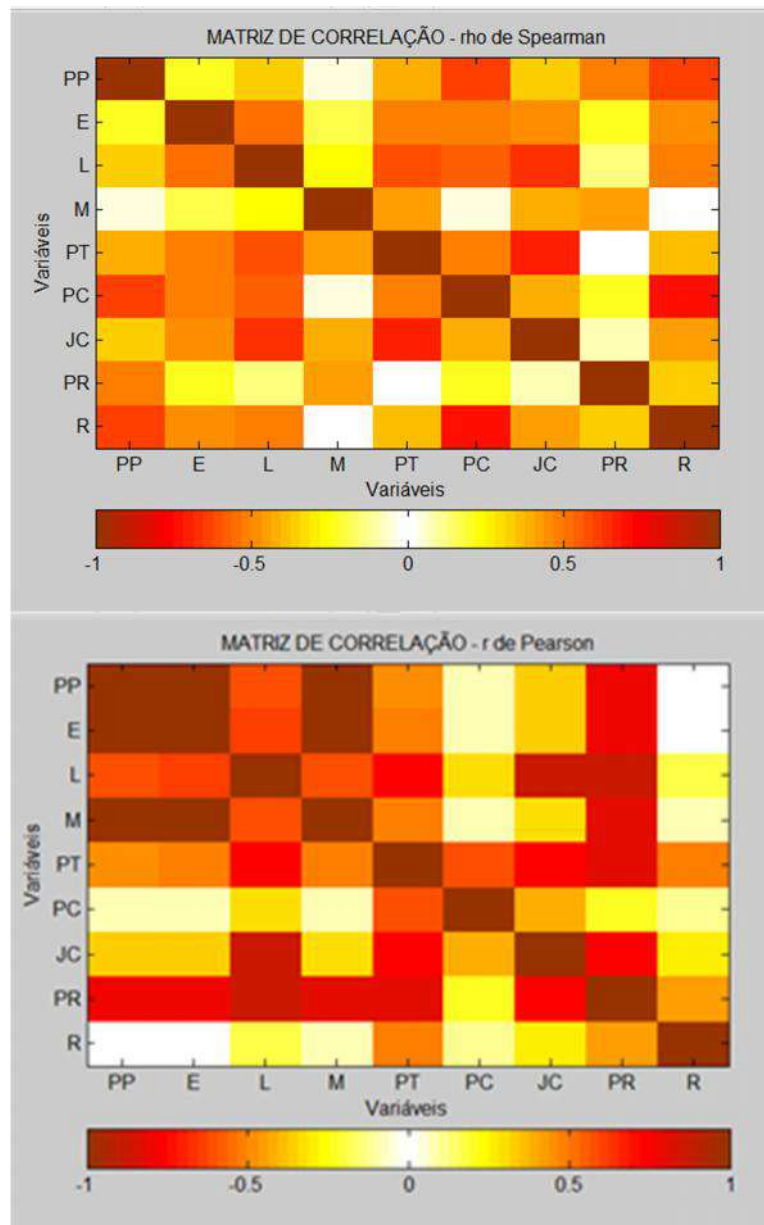


Figura 37 – Coeficiente de Assimetria para as estações de monitoramento de qualidade de água.

As Figuras 38 e 39 apresentam a matriz de correlação para os coeficientes r de Pearson e Rho de Spearman respectivamente. É possível observar que a correlação linear e a associação para os 36 possíveis pares de estações (ou seja, combinações de 9 locais de monitoramento tomadas 2 de cada vez), o coeficiente r de Pearson e o coeficiente de Spearman mostram uma forte correlação (ou seja, $|r| > 0.5$ or $|p| > 0.5$)

para 45% dos pares para r e todos os pares para ρ ; além de o coeficiente r de Pearson mostrar correlação moderada (isto é, $0.3 < |r| < 0.5$) para 25% dos pares. Nota-se que, para ambos os coeficientes, r e ρ , a correlação linear e a associação entre as estações foram maiores do que entre os parâmetros da qualidade da água. Estes resultados estão de acordo com os encontrados por Ouyang (2005) que afirma que a relação entre as estações é mais estática por ser controlada por condições hidrogeológicas. Assim como na aplicação anterior, foram observadas distribuições assimétricas das variáveis e foi aplicada uma transformação logarítmica antes da ACP.



Figuras 38 e 39 - Matriz de correlação para as estações de monitoramento de qualidade de água com base nos coeficientes r de Pearson e ρ de Spearman respectivamente.

Considerando a *Kaiser's rule*, dois componentes apresentaram autovalores superiores a um (Tabela 10). Por outro lado, os dois primeiros componentes explicaram apenas 65% da variabilidade observada, que é inferior aos 70% exigidos pelo critério Porcentagem de Variação Total (Tabela 11). Em contrapartida, o gráfico de *scree test* mostrou que mais de dois componentes deviam ser considerados (Figura 40). Desta forma, foram selecionados três componentes, que representaram 75% da variância total. Esses resultados também estão de acordo com Ouyang (2005), que apontou a necessidade de usar menos componentes para locais de monitoramento do que parâmetros de qualidade da água. As estações mais relevantes foram Morin (M) e Esperança (E) (valores dos *loadings* apresentados em negrito na Tabela 11).

Tabela 10- Os autovalores e variâncias do PCA para os locais de monitoramento.

Componentes	Autovalores	Variância (%)
1	2.89	45.15
2	1.28	20.08
3	0.65	10.22
4	0.56	8.78
5	0.32	5.00
6	0.30	4.75
7	0.20	3.14
8	0.18	2.89
9	0.00	0.00

Tabela 11 - Os *loadings* dos três componentes principais após a rotação Varimax para os locais de monitoramento.

Estações/Componentes	Loadings		
	1	2	3
Parque Petropolis (PP)	0.14	-0.56	0.23
Esperança (E)	0.16	0.15	0.82
Liceu (L)	0.50	-0.10	0.51
Morin (M)	0.97	0.38	0.02
Poço Tarzan (PT)	-0.62	0.13	-0.19
Poço do Casinho (PC)	-0.40	0.34	-0.53
Joao Christ (JC)	-0.52	0.08	-0.26
Pedro do Rio (PR)	-0.10	-0.64	-0.06
Rocio 2 (R)	-0.13	0.22	-0.55

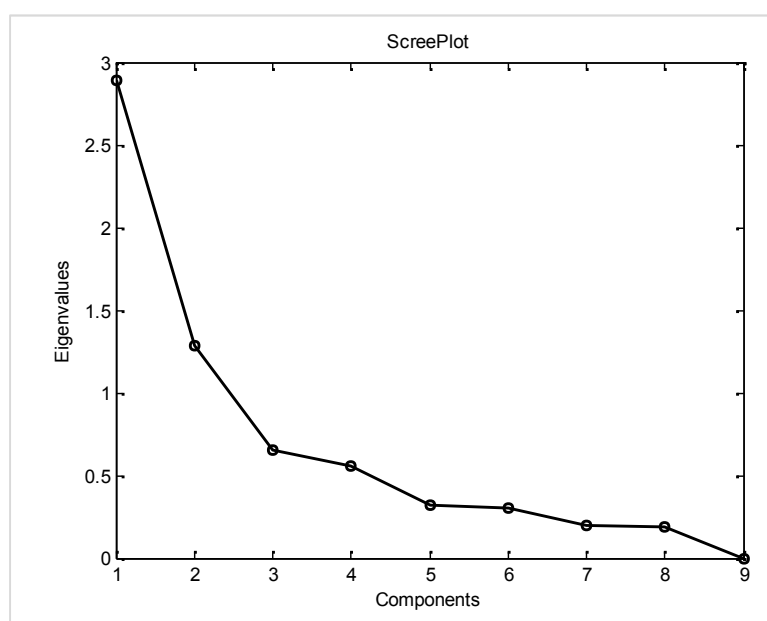


Figura 40 - O *scree plot* após aplicação do PCA aos locais de monitoramento.

Para ACPNL/RNAA/IG, várias arquiteturas de rede foram testadas com valores crescentes de f a partir de um, e valores de $M1$ e $M2$ começando em dois e aumentando a cada dois. A opção com dois componentes principais ($f=2$) e quatro neurônios nas camadas de mapeamento/desmapeamento teve o melhor desempenho baseado em EMQ, EPF e CTI (Tabela 12). O IG de cada variável é mostrado na Tabela 13.

Tabela 12 - Principais arquiteturas de redes neurais utilizadas na ACPNL/RNAA/IG para avaliação dos locais de monitoramento.

	f=1			f=2			f=3		
M1/M2	EMQ	EPF	CTI	EMQ	EPF	CTI	EMQ	EPF	CTI
6	0.074	0.007	-7.371	0.045	0.004	-7.589	0.055	0.005	-6.471
4	0.071	0.006	-7.510	0.035	0.003	-7.959	0.036	0.003	-7.378
2	0.035	0.003	-7.540	0.040	0.003	-7.061	0.036	0.003	-6.990

Tabela 13 - O IG (%) calculado para os locais de monitoramento.

Estações de Monitoramento	IG (%)
Rocio 2 (R)	18
Liceu (L)	15
ParquePetropolis (PP)	12
Poço do Casinho (PC)	11
Poço Tarzan (PT)	11
Morin (M)	10
Joao Christ (JC)	10
Pedro do Rio (PR)	7
Esperança (E)	4

4.2.5 Discussão

Parâmetros de Qualidade de Água

É possível observar a partir dos resultados apresentados que cada componente na ACP representa uma classe de parâmetros de qualidade da água: o primeiro e o segundo (ou seja, NH_4^+ , DQO e OD para o primeiro e Al para o segundo) estão relacionados a parâmetros químicos; o terceiro (ou seja, Turb), para parâmetros físicos; e o quarto (isto é, CF e CT), para parâmetros biológicos (Tabela 7). Estes resultados estão alinhados com os resultados encontrados por SIMEONOV *et al.* (2003), PHUNG *et al.* (2015) e OGWUELEKA (2015). Além disso, NH_4^+ , OD, CF, Turb e CT, que são parâmetros relevantes, provavelmente estão relacionados à poluição causada por esgoto doméstico não tratado, definitivamente uma das principais questões na área urbana (ver valores na Tabela 3). Além disso, a presença de Al como um dos parâmetros significativos e sua

alta concentração na área agrícola pode refletir o uso de sulfato de alumínio na tentativa de se aumentar a acidez do solo nas áreas de plantio.

A partir dos resultados da ACPNL/RNAA/IG, apresentados na Tabela 7, é possível perceber que a relevância está bem distribuída entre o conjunto de parâmetros selecionados para o estudo e é difícil separar um conjunto de parâmetros mais relevantes e menos relevantes. Provavelmente este conjunto de parâmetros selecionado seja adequado para caracterizar a variabilidade da qualidade da água na área em estudo. RESINO *et al.* (2011) encontraram resultados similares ao aplicarem a técnica IG para determinar a relevância de 10 variáveis no intuito de prever características de fibrose avançada. Ainda relacionado a esse estudo, todos os valores IG estavam próximos uns dos outros, variando entre 4% e 14%. Os autores concluíram que todas as variáveis tinham uma contribuição relevante para a variável de saída. Ao mesmo tempo, o IG do parâmetro mais relevante, Coliformes Fecais (CF), é três vezes o valor dos de menor relevância, Turbidez (Turb) e Demanda Química de Oxigênio (DQO). O índice de coliformes fecais está relacionado à poluição das águas residuais domésticas e, conseqüentemente, sua concentração é muito maior na bacia urbana (Tabela 5). Seguindo a ordem de relevância, os seguintes cinco parâmetros, Nitrato (NO_3), Temperatura (Temp), Oxigênio Dissolvido (OD), Sulfato (SO_4^{2-}), e Ferro (Fe), têm valores IG maiores do que o dobro do parâmetro menos relevante. O nitrato tem concentrações mais elevadas na bacia agrícola, o que sugere que este índice poderia estar associado ao uso de fertilizantes. Isso está de acordo com as conclusões de VEGA *et al.* (1998), WANG *et al.* (2012), OGWUELEKA (2015) e TANOS *et al.* (2015), que apontam a presença do nitrato na água como resultado do uso de fertilizantes. Com base nesse resultado, é possível demonstrar a importância da bacia hidrográfica experimental agrícola na variabilidade da qualidade da água. Temperatura e oxigênio dissolvido são medidos com uma sonda multiparamétrica, com menores custos de monitoramento, o que é importante no caso de restrições orçamentárias. Temperatura e oxigênio dissolvido também são indicadores da variabilidade temporal/sazonal da qualidade da água (OUYANG, 2006; OGWUELEKA, 2015). As fontes de Ferro e Sulfato podem ser provenientes da lixiviação do solo ou das águas residuais industriais/domésticas (SIMEONOV *et al.*, 2003, SELLE *et al.*, 2013). A segunda opção parece mais razoável, dado o seu alto valor médio na bacia urbana (Tabela 5). A partir desta análise, pode-se observar que o RMQAP está monitorando o impacto dos diferentes usos do solo na

qualidade da água, alcançando, assim, o objetivo para o qual foi planejada. Além disso, há uma predominância de parâmetros relacionados à área urbana, conforme encontrado na aplicação ACP.

Comparando-se as variáveis mais relevantes das diferentes técnicas ACP e ACPNL/RNAA/IG percebe-se que dois parâmetros estavam presentes em ambos os métodos, são eles: Oxigênio Dissolvido e Coliformes Fecais. Tais parâmetros estão relacionados à poluição orgânica, que é a principal problemática da bacia hidrográfica. Provavelmente, eles têm um comportamento predominantemente linear. Por outro lado, a Demanda Química de Oxigênio e a Turbidez, que foram os parâmetros menos relevantes na utilização da ACPNL/RNAA/IG, foram considerados relevantes pela ACP. Além disso, a ACPNL/RNAA/IG explica cerca de 80% da variância do conjunto de dados, obtendo resultados ligeiramente superiores do que com o ACP, que explica aproximadamente 76%. Embora a diferença nos percentuais de variância não tenha sido grande, o desempenho da ACPNL/RNAA/IG para reconstruir o conjunto de dados, por outro lado, foi muito superior do que ACP com base em EMQ, EPF e CTI e R^2 , em uma dimensão inferior apresentados na Tabela 14. Em outras palavras, usando-se um menor número de componentes, a ACPNL/RNAA/IG reproduziu os dados de qualidade da água com uma precisão superior ao ACP (Tabela 12).

Tabela 14. Os resultados dos métodos ACP e ACPNL/RNAA/IG para os parâmetros de qualidade de água

Modelos	EMQ	EPF	AIC	R^2
ACPNL/RNAA/IG	0.02	0.02	-3.91	0.97
ACP	0.24	0.11	-2.18	0.87

Estações de Monitoramento

Com base no maior valor absoluto *loadings PCA* e do seu respectivo componente, as estações poderiam ser agrupadas de acordo com sua localização na bacia hidrográfica (Tabela 11). Nota-se, que as estações na bacia agrícola (PT, JC e PC) e na bacia preservada (R) contribuem negativamente para o primeiro e terceiro componentes; enquanto que as estações localizadas na área urbana (E, L e M) contribuem positivamente para o primeiro e terceiro componentes. O segundo

componente, por outro lado, foi amplamente influenciado pelas estações a jusante das três bacias hidrográficas experimentais (PP e PR).

Observa-se na Tabela 13 que, embora a relevância esteja bem distribuída entre as variáveis de entrada, é possível separar as duas estações mais relevantes, Rocio (*R*) e Liceu (*L*) e as duas menos relevantes Pedro do Rio (*PR*) e Esperança (*E*), de um “bloco” de estações com valores IG muito similares (aproximadamente 11%). As técnicas, ACP e ACPNL/RNAA/IG apresentaram resultados relativamente contraditórios. Enquanto no ACP foram consideradas relevantes duas estações da bacia urbana, a ACPNL/RNAA/IG indicou uma estação na bacia preservada e outra na bacia urbana. O “bloco” de estações é dominado pelas estações da bacia agrícola. Por outro lado, uma das estações menos relevantes localiza-se área urbana (*E*) indicando que provavelmente ela não é representativa e deva ser removida ou reposicionada. Assim, mostrando que parte da rede não está funcionando de acordo com o objetivo do RMQAP, que é avaliar o impacto de diferentes usos na bacia. Por outro lado, a estação de Esperança foi considerada relevante para PCA. A estação de Pedro do Rio está localizada longe das bacias experimentais e, portanto, esperava estar na base da lista do "ranking de estações". MORIASI (2015) afirma que as estações da RMQAP localizadas bem próximas geograficamente, ou seja, na mesma bacia experimental, geram informações redundantes e sugere um possível remanejamento das estações.

A *FEV* da ACP foi de 75%, enquanto que da ACPNL/RNAA/IG foi de 96%, usando um número menor de componentes. Neste caso, houve uma melhoria substancial da variância explicada mostrando maior capacidade de representação dos dados pela metodologia proposta. A Tabela 15 mostra uma diferença significativa entre os desempenhos na reconstrução de dados dos dois métodos. A ACPNL/RNAA/IG teve resultados muito melhores considerando os quatro critérios de avaliação (EMQ, EPF e CTI e R^2), assim como no caso anterior.

Tabela 15. Os resultados dos métodos ACP e ACPNL/RNAA/IG para os locais de monitoramento.

Modelos	EMQ	EPF	AIC	R^2
ACPNL/RNAA/IG	0.00	0.00	-6.77	0.98
ACP	0.36	0.19	-1.67	0.73

Em suma, é possível observar diferenças expressivas entre os resultados dos dois métodos. Provavelmente, as variáveis de entrada (isto é, dados de qualidade da água) possuem um comportamento não linear significativo. Caso contrário, os resultados deveriam ter sido bastante semelhantes, uma vez que a ACPNL é capaz de avaliar relacionamentos lineares e não-lineares. Ao mesmo tempo, em ambos os casos, a ACPNL/RNAA/IG apresentou melhor desempenho com base na avaliação da variância e na capacidade de reconstrução de dados. Estes resultados estão de acordo com os encontrados por MONAHAN (2000) E RAZAVI *et* COULIBAL (2013). Na ACP foi selecionado um maior número de componentes principais do que ACPNL/RNAA/IG para explicar uma menor parte da variação de dados em ambos os casos avaliados. Isso significa que os componentes selecionados no ACPNL/RNAA/IG representam muito melhor os dados originais. Além disso, reflete a maior variabilidade da qualidade da água, uma vez que sacrifica uma menor porcentagem da variância dos dados. O que está de acordo com THEODORIS *et* KOUTROUMBAS (2009) que afirmam que quando os dados são altamente não-lineares, a ACP tende a superestimar o número de componentes.

Durante muito tempo, a análise de componentes principais (em suas formas lineares e não-lineares) ficou restrita a áreas de conhecimento muito específicas devido à dificuldade de se aplicar seus conceitos matemáticos sofisticados. No entanto, os avanços tecnológicos em computação nas últimas décadas fizeram com que usuários sem o conhecimento aprofundado do método se beneficiassem de suas múltiplas funcionalidades (JOLLIFFE, 2002). Software comerciais e gratuitos têm favorecido a manipulação e análise de dados trazendo o ACP para órgãos do governo, indústria e universidades (HAIR *et al.*, 2009). Diferentes áreas, como, por exemplo, a área de monitoramento da qualidade da água começou a usar o ACP devido à urgente necessidade de ferramentas automatizadas para auxiliar os usuários a transformar a grande quantidade de dados produzidos em informações úteis (HAN *et al.*, 2011a). Desta forma, usuários com diferentes níveis de conhecimento podem aplicar a metodologia proposta com auxílio de programas computacionais e de uma interface amigável. Tal interface é importante para superar as limitações da compreensão da metodologia e possibilitar o alcance de um maior número de pessoas que poderiam se beneficiar de seu uso, das quais cabe citar os gestores de redes de monitoramento. Assim, estaria disponível para eles um ranking simples com porcentagens de relevância

de cada parâmetro de monitoramento e das estações. Esse ranking seria útil para definir prioridades de gerenciamento, tanto no caso de investimento com no caso de restrições orçamentárias custos. Com base nos resultados apresentados, o gestor da RMQAP na CPRM/SGP poderia, por exemplo, em caso de cortes de despesas, remover a estação Esperança, interromper ou reduzir a frequência do monitoramento de DQO ou continuar monitorando temperatura e oxigênio dissolvido porque são relevantes e acessíveis por exemplo. Alternativamente, seria possível realocar a estação Esperança ou investir na análise de coliformes fecais, o parâmetro mais relevante, para obter dados mais precisos, no caso de investimento financeiro na robustez da rede.

5 Avaliação da RMQAP com base em modelagem hidrológica e redes neurais

Nesse capítulo será apresentada a segunda metodologia proposta para avaliação da RMQAP com base em modelagem hidrológica com o uso do SWAT a partir do desenvolvimento de índice de mensuração de impacto da estação na calibração do modelo relacionado às redes neurais;

5.1 Metodologia

A metodologia proposta tem como base o uso de modelagem hidrológica e de redes neurais para a simulação da qualidade de água na área de estudo e, posterior, avaliação da RMQAP. De acordo com seguinte pressuposto: se existe um modelo capaz de representar a bacia razoavelmente, é possível avaliar a sua rede de monitoramento com base no impacto gerado pelos dados da estação na calibração do modelo, ou seja, na sua acurácia.

Dessa forma, por exemplo, é possível realocar ou eliminar estações onde o modelo tenha uma boa representação e os dados observados não melhorem o seu ajuste ou, então, reduzir a frequência de monitoramento nos locais da bacia onde o modelo represente com boa acurácia. Em outras palavras, a avaliação da RMQA se dá em função da análise do desempenho do modelo em relação aos dados das estações de monitoramento.

O impacto de cada estação no ajuste do modelo é composto por dois tipos de impacto: o impacto dos dados da estação na calibração de vazões e pelo impacto dos dados da estação na calibração de cada parâmetro de qualidade de água. Isso porque a calibração de qualidade de água requer a calibração de vazões no modelo SWAT. Além disso, a calibração de qualidade de água, geralmente, é bem mais complexa e complicada que a de vazões. E, ainda, há poucos estudos que obtêm resultados adequados, quando comparados aos de vazão (ARNOLD et al., 2012).

O peso do impacto de cada tipo de calibração na estimativa do impacto final da estação é obtido a partir do ajuste de um modelo de redes neurais para a previsão dos dados de qualidade de água em função dos dados de vazão da estação. As redes neurais possuem um ótimo potencial para esse tipo de previsão e inclusive já foram utilizados

para este fim (BOWERS *et al.* 2000; SARKAR *et al.*, 2010; MAIER *et al.*, 2010). Isso porque se é possível fazer essa previsão com certa acurácia, significa que os dados de vazão, mais facilmente obtidos, devem ser priorizados em função dos de qualidade de água mais custosos na minha rede de monitoramento.

A metodologia requer a definição de cenários de calibração a partir do estabelecimento de grupamento de estações da rede que se pretende avaliar de acordo com o objetivo ou finalidade da avaliação da rede. No caso, pretende-se avaliar se a RMQA está alcançando os objetivos para os quais ela foi planejada. No caso da RMQAP, o objetivo principal é a avaliação do impacto dos usos agrícola e ocupação urbana na qualidade da água com o menor custo possível, já que o projeto EIBEX sofre constantes restrições orçamentárias.

Nesse caso, é definido um cenário otimista que utiliza apenas a estação do exutório da bacia, representando o limite máximo de redução de custos; e um cenário pessimista que engloba todas as estações e representa o limite mínimo de redução de custo, isto é, a situação atual da rede. Além desses cenários, deve-se definir um cenário “padrão” que corresponde à saída do modelo sem que seja realizado nenhum ajuste de parâmetros. Esse cenário equivale à representação do modelo sem levar em conta os dados da rede de monitoramento. Ele é resultado apenas da representação física da bacia tomando por base os dados de entrada do modelo. E, por último deve-se definir o cenário que será utilizado como base de comparação para os demais para a mensuração do impacto dos dados da estação na calibração. Nesse cenário cada estação é calibrada individualmente, sem interferência das outras estações, representando o impacto individual de cada uma de forma isolada e seu potencial máximo de impacto na calibração do modelo.

Assim, para a estimativa do impacto de cada estação na RMQAP com base na metodologia proposta, foi estabelecido um índice de avaliação da RMQAP (I_{RMQAP}), expressado pela Eq. 12. O índice deve ser usado para o cálculo o impacto de cada parâmetro de qualidade de água separadamente.

$$I_{RMQAP}^e = AI^e \left(F1_{IND}^e \frac{NSE_Q^e}{\sum_{i=0}^n ER_Q^e} + F1_{IND}^e \frac{NSE_{QA}^e}{\sum_{i=0}^n ER_{QA}^e} \right) \quad (12)$$

onde, e é a estação que se pretende avaliar; AI é a área incremental percentual de contribuição da estação e . AI é calculada a partir da soma das áreas de contribuição das sub-bacias a montante da estação, nos casos das estações de cabeceira, ou da área de contribuição das sub-bacias entre a estação e a última estação a montante, dividida pela área total da bacia. NSE_Q^e e NSE_{QA}^e são os coeficientes de eficiência Nash e Sutcliffe (NSE) para avaliação do ajuste do modelo na calibração das vazões (Q) e do parâmetro de qualidade de água (QA) respectivamente da estação e . NSE é calculado a partir da Eq. 13 e determina a magnitude relativa da variância residual em comparação com a variação dos dados medidos e é um das métricas estatísticas mais utilizadas para avaliação da calibração do modelo SWAT (MORIASI *et al.*, 2007). O NSE varia de - a 1 sendo esse seu valor ótimo, valores negativos significam que a média dos valores observados é melhor preditor que os valores simulados. Por essa razão, para os casos de valores negativos de NSE, estabeleceu-se que o coeficiente de ponderação será igual zero, e por consequência, o impacto daquele tipo de calibração.

$$NSE = 1 - \frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^n (Y_i^{obs} - Y^{med})^2}$$

(13) onde Y_i^{obs} é a i^a observação da variável observada; Y_i^{sim} é a i^a observação da variável simulada; Y^{med} é a média dos dados observados e n o número total de observações.

$F1^e$ e $F2^e$ são os fatores de ponderação ou pesos utilizados para ponderar o impacto da calibração de vazões e da calibração do parâmetro de qualidade em relação ao impacto final da estação e . Referem-se ao ajuste das redes neurais para a previsão dos dados de qualidade de água em cada estação individualmente. $F1$ é mensurado a partir da média das métricas estatísticas: NSE e coeficiente de Determinação (R^2) para o cenário IND que representa o impacto individual da estação e . ER_Q^e e ER_{QA}^e são os erros relativos percentuais entre o cenário padrão (cenário 0) e o cenário n para a calibração de vazões e de qualidade de água respectivamente, sendo n o número de cenários. Os erros relativos devem ser padronizados entre os valores 0 e 1 para cada cenário. O erro oferece uma estimativa do prejuízo que a calibração da estação sofre com o ajuste da calibração naquele cenário. Ou seja, quanto maior o erro maior a estação sofre com o ajuste do cenário em relação ao seu potencial máximo estabelecido no cenário 0.

Objetivamente, a metodologia consiste nas seguintes etapas:

- Seleção dos dados;
- Definição de cenários de rede de monitoramento de acordo com o objetivo da avaliação;
- Simulação e avaliação dos cenários para vazões e qualidade de água a partir das métricas estatísticas usando o modelo SWAT;
- Simulação dos dados de vazão para previsão dos dados de qualidade de água usando redes neurais;
- Estimativa do impacto de cada estação a partir do cálculo do I_{RMQA}^e ;
- Avaliação da RMQA.

Na etapa de seleção de dados é realizada a verificação da compatibilidade dos dados de monitoramento com o modelo utilizado. Nessa fase são avaliadas as limitações do modelo e realizadas atividades preparatórias dos dados.

O próximo passo consiste na simulação dos cenários utilizando o modelo definido, onde estão incluídos os processos de calibração e validação. Além dos cenários pré-estabelecidos é importante que se realize a calibração para cada estação isoladamente.

Os cenários serão calibrados usando apenas as estações daquele cenário. Isso quer dizer que para cada cenário os parâmetros do modelo serão ajustados a partir da comparação dos dados observados da(s) estação (ões) que compõe (m) o cenário com os dados simulados. Em seguida, o modelo ajustado será aplicado à rede completa, ou seja, as demais estações que não integram o cenário. Esse procedimento seria uma espécie de validação espacial do modelo calibrado (ARNOLD *et al.*, 2012). A seguir, o valor simulado nas estações que não compõem o cenário será comparado com os valores observados e as métricas técnicas estatísticas calculadas.

A avaliação da RMQA deve ser feita com base no I_{RMQA}^e associado aos demais aspectos relativos ao processo de simulação do modelo SWAT e das redes neurais.

5.2 Estudo de Caso

A aplicação da metodologia foi feita para a bacia representativa do Rio Piabanha tendo como base a rede de monitoramento hidrometeorológica operada pelo SGB/CPRM de acordo com o passo-a-passo estabelecido na seção 5.1. Serão utilizados dois modelos, primeiramente, o modelo hidrológico SWAT, em seguida, as Redes Neurais para que ao final os resultados sejam comparados entre si.

5.2.1 Modelagem Hidrológica - SWAT

A modelagem hidrológica a partir do modelo SWAT será realizada tendo como base etapas da metodologia proposta,

5.2.1.1 Seleção do Dados

Os dados usados no processo de calibração tem interferência direta na validação e avaliação dos resultados da modelagem hidrológica (MORIASI *et al.*, 2007). Normalmente, recomenda-se um período de 10 anos de dados para a calibração, mas na falta de dados disponíveis, uma prática comum é a separação de 2/3 da série de dados para a calibração e 1/3 para validação (BRESSIANI, 2016). MORIASI *et al* (2007) afirma que a calibração ideal deve usar entre 3 a 5 anos de dados incluindo períodos anos secos e úmidos e uma gama de eventos hidrológicos suficientes. Como já relatado anteriormente, a rede hidrometeorológica da bacia representativa do rio Piabanha começou a operar em 2007 mas os dados iniciais estavam com muitas falhas e inconsistências. Optou-se em usar a serie toda para calibração no período de 2009 a 2012, sendo os anos 2007 e 2008 usados para aquecimento do modelo.

Já para as simulações de qualidade de água, devido a pouca disponibilidade de dados usou-se todo o conjunto para a calibração relativo ao período de 2009-2012, já que a RMQAP começou a operar em 2009.

Foram selecionadas para a calibração as mesmas estações de monitoramento utilizadas na aplicação da primeira metodologia apresentadas na Tabela 2 e Figura 31.

Basicamente, o SWAT requer quatro tipos de informações de entrada referentes à área de estudo que devem ser fornecidos pelo usuário:

- a. O modelo digital de elevação (MDE);

O modelo digital de elevação foi fornecido a partir do SRTM, inicialmente, com resolução de 90m, que foi substituído pelo de 30m conforme detalhado na seção 3.4.1.

- b. Informações sobre uso e ocupação do solo;

Foram utilizados dois mapas de uso e ocupação do solo, conforme detalhado na seção 3.4.3, um na escala 1: 250.000 e outro na escala 1:25.000, ambos projetados em UTM-WGS_84- 23S.

- c. Os tipos de solo;

As informações sobre os tipos de foram reunidas com base no mapa elaborado pelo levantamento elaborado pela Embrapa e detalhado na seção 3.4.2. O relatório do levantamento de solos traz uma caracterização das classes de solo dominantes, acompanhada da descrição e resultados analíticos de alguns perfis representativos, assim como uma descrição geral dos padrões característicos de distribuição dos solos no estado. Entretanto, o documento não fornece informações detalhadas das classes de solo necessárias para diversos estudos de modelagem hidrológica. O fato dos solos serem classificados em associações ao invés de uma única classe também dificulta a obtenção de uma distribuição espacial mais fidedigna da região. Dessa forma, para o desenvolvimento desse trabalho procedeu-se as seguintes definições:

1) Classe de Solo dominante – como citado anteriormente, para a bacia representativa foram observadas nove associações de, basicamente, quatro classes de solo. No relatório, são definidos os percentuais de cada classe, sendo apresentada, primeiramente, a classe de maior área percentual seguida das demais em ordem decrescente. Como para esse estudo era preciso definir a distribuição espacial de cada classe de solo, e não foram encontrados estudos mais detalhados capazes de identificar espacialmente às porções, separadamente, optou-se em considerar a classe dominante em toda área da respectiva associação, com exceção, das unidades AR3 e AR5, cuja classe dominante era afloramento rochoso. Nesse caso, como o mapa de uso e ocupação também a classe “afloramento rochoso”, e no fato de que ele será usado, também, nas definições de entrada para a modelagem, optou-se em usar a classe de solo seguinte para as respectivas áreas. A Tabela 16 apresenta a classe dominante usada para cada associação ou unidade de solo, o perfil utilizado (Embrapa, 2003) e o grupo hidrológico definido.

2) Definição das características da classe – para o estudo aqui proposto eram necessárias informações detalhadas de cada tipo do solo extraídas, normalmente, dos perfis e amostras de solo, tais como: a composição granulométrica, o teor de carbono orgânico, textura, capacidade de água disponível, etc. Dessa forma, como não foram apresentados os perfis das classes de solo referentes à área de estudo, foram selecionados os mais próximos e extraídas as informações sobre o teor de carbono orgânico, composição granulométrica e profundidade de cada horizonte do solo. Com essas informações e o auxílio de uma macro desenvolvida por SAXTON et RAWLS (2006), foram calculados as demais características necessárias para a modelagem.

3) Grupos Hidrológicos – O Serviço de Conservação dos Recursos Naturais americano elaborou um sistema de classificação de solos, em quatro grupos hidrológicos (A, B, C e D) segundo suas características de infiltração (NRCS, 2007). Essa informação, também, era necessária para a modelagem. Assim, com base nas características de cada classe apresentadas no Relatório (Embrapa, 2003) e no artigo de SARTORI et al. (2005) foi realizada a classificação.

Tabela 16 – Definição das características das classes do solo da bacia representativa.

Unidade/ Associação	Classes de Solo (EMBRAPA, 2003)	Classe Predominante	Perfil de solo (EMBRAPA, 2003)	Grupo Hidrológico (NRCS, 2007)
AR3	AFLOREAMENTOS DE ROCHA + CAMBISSOLO HÁPLICO Tb + NEOSSOLO LITÓLICO Distrófico (50	Cambissolo	11	C
AR5	AFLOREAMENTOS DE ROCHA+ NEOSSOLO LITÓLICO Distrófico (70-30%)	Neossolo	29	D
CXbd12	CAMBISSOLO Álico Tb + AFLOREAMENTOS DE ROCHA + LATOSSOLO VERMELHO-AMARELO Álico (50-30-20%)	Cambissolo	10	C
CXbd4	CAMBISSOLO Tb + LATOSSOLO VERMELHO- AMARELO (80-20%)	Cambissolo	10	C
CXbd6	CAMBISSOLO Tb + LATOSSOLO VERMELHO- AMARELO + LATOSSOLO VERMELHO-AMARELO (50-30-20%)	Cambissolo	10	C
CXbd8	CAMBISSOLO Tb + LATOSSOLO VERMELHO- AMARELO + LATOSSOLO VERMELHO-AMARELO (50-30-20%)	Cambissolo	11	C
LVAAd24	LATOSSOLO VERMELHO-AMARELO + CAMBISSOLO Tb t (80-20%)	Latossolo	23	A
LVAAd26	LATOSSOLO VERMELHO-AMARELO + CAMBISSOLO Tb (70-30%)	Latossolo	23	A
RLd	SOLOS LITÓLICOS (NEOSSOLOS) + CAMBISSOLO Álicos Tb A (80-20%)	Neossolo	29	D

d. Dados climáticos.

O SWAT utiliza as cinco variáveis climáticas: precipitação, temperatura máxima e mínima, radiação solar, velocidade do vento e umidade relativa. Assim, foram reunidos dados diários de estações pluviométricas e climatológicas de diferentes redes de monitoramento: Instituto Nacional de Meteorologia (INMET, 2013), The National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (NCEP/CFSR, 2013), Rede Nacional Hidrometeorológica (RNH, 2013) e da RMQAP. A Tabela 17 apresenta as informações das estações selecionadas para o estudo e a Figura 41. É possível observar que as estações possuem período de dados diferentes. Assim, como SWAT requer o mesmo período de dados para todas as estações, optou-se em preencher o período faltante utilizando Redes Neurais através de um estudo de correlação entre as estações onde foi comparado o desempenho da regressão e das redes neurais.

Tabela 17 – Estações selecionadas para entrada de dados.

Código	Tipo	Latitude (graus dec)	Longitude (graus dec)	Elevação (m)	Rede	Período (anos)
A603	C	-22.583	-43.267	33	INMET	2002-2012
A610	C	-22.450	-43.283	1777	INMET	2007-2012
A618	C	-22.433	-42.933	980	INMET	2007-2012
223431	C	-22.324	-43.125	830	NCEP/CFSR	1980-2010
2243286	C	-22.405	-43.133	778	RMQAP	2007-2012
2243287	P	-22.511	-43.210	819	RMQAP	2007-2012
2243289	P	-22.487	-43.177	815	RMQAP	2007-2012
2243288	P	-22.517	-43.169	818	RMQAP	2007-2012
2243301c	P	-22.461	-43.095	1004	RMQAP	2007-2012
2243291	P	-22.469	-43.103	1030	RMQAP	2007-2012
2243302	P	-22.478	-43.258	972	RMQAP	2007-2012
2243012	P	-22.332	-43.134	654	RMQAP/RNH	1938-2012
2243010	P	-22.485	-43.149	1085	RNH	1938-2012
2243011	P	-22.438	-43.170	704	RNH	1938-2012
2243014	P	-22.300	-43.178	460	RNH	1938-2012
2243268	P	-22.511	-43.210	819	RNH	2005-2012

Além disso, o SWAT possui um modelo de geração de dados climáticos que calcula métricas estatísticas, necessárias para preenchimento de falhas e para a geração de dados sintéticos, denominado *WXGEN* (NEITSCH *et al*, 2009). Para isso, devem ser

fornechos dados das mesmas cinco variáveis climáticas, listadas anteriormente. São recomendados, idealmente, 20 anos de dados e as estações utilizadas podem ter períodos diferentes de dados (NEITSCH *et al*, 2009). Nesse caso, foram selecionadas quatro estações climatológicas para a geração desses dados: A618, A610 e A603, do INMET e 2243286, da RMQAP.

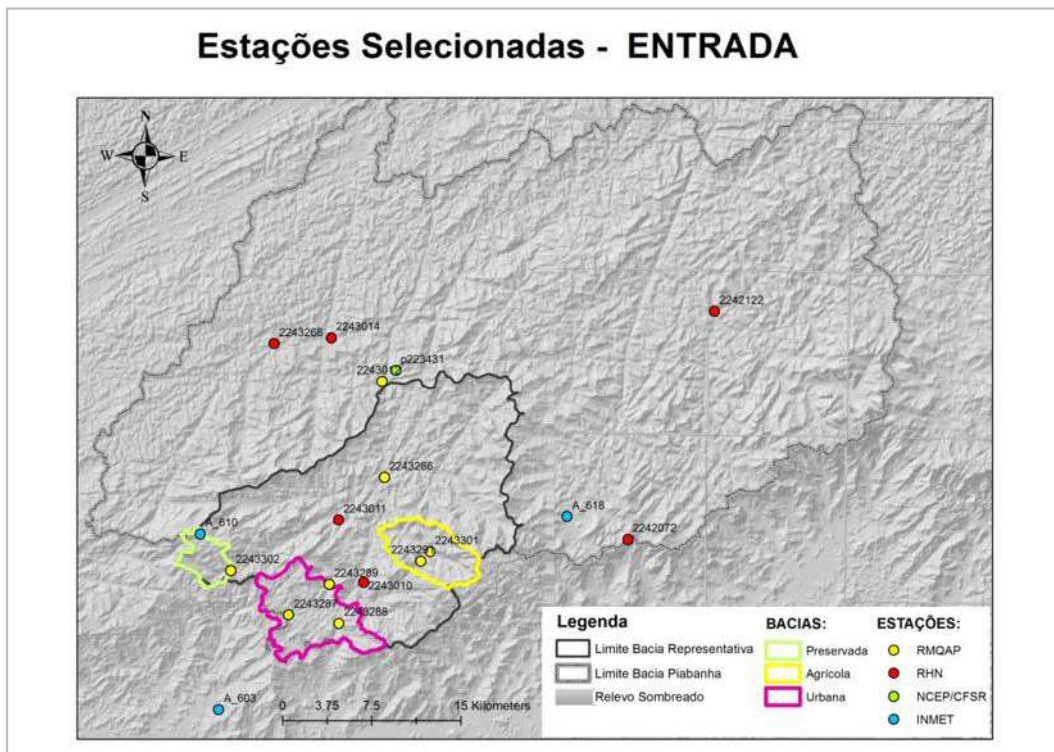


Figura 41 – Estações selecionadas para obtenção dos dados climáticos de entrada.

e. Dados de vazão e qualidade de água.

Foram selecionadas as séries de vazões diárias das nove estações fluviométricas que coincidem com as de qualidade de água e compõem a RMQAP (Tabela 2). As séries foram determinadas a partir de dados convencionais (*i.e.* 2 leituras de réguas diárias por observador hidrológico) e passaram por uma avaliação preliminar para retirada de inconsistências e correção de algumas falhas. A partir dos dados diários foram produzidas as séries mensais para cada estação. As Figuras 44, 41 e 43 apresentam as séries diárias de vazão para cada estação selecionada organizada por bacia experimental e a Figura 45 para as estações parque Petrópolis e Pedro do Rio. Nota-se que na bacia rural, as estações Poço Tarzan e João Christ apresentam muitas falhas, a maior parte deles, por problemas na medição com a régua que é levada com a

cheia. Os cursos d'água dessa área da bacia apresentam altas declividades com vazões de cheias muito elevadas conhecidas popularmente como “cabeças d'água”. Os níveis se elevam rapidamente e demoram pouco tempo para diminuir. Dessa forma, para uma melhor caracterização das vazões da bacia o ideal seriam medições com frequência horária ou sub-horária. O pico de vazão em janeiro de 2008 aparentemente parece inconsistente, mas ele ocorreu nas três estações e inclusive danificou a régua da estação Poço do Casinho. A estação da bacia preservada foi instalada mais tardiamente, por isso a série de dados é mais curta. Aqui também observa-se altas declividades dos cursos d'água e a problemática de danificação dos equipamentos em virtude das “cabeças d'água” (por exemplo, o pico de vazão em janeiro de 2011 seguido por período de falhas). Embora tenha uma série menor de dados esta estação apresenta poucas falhas quando comparada as da bacia rural. As estações da bacia urbana formam o conjunto com o menor percentual de falhas. Entretanto, nota-se diversos “patamares” nos dados de Liceu que podem representar leituras duvidosas do observador. Por outro lado, a estação Pedro do Rio apresenta, aparentemente, uma série consistente com pouquíssimas falhas. Já a estação Parque Petrópolis, apresenta algumas lacunas na sua série.

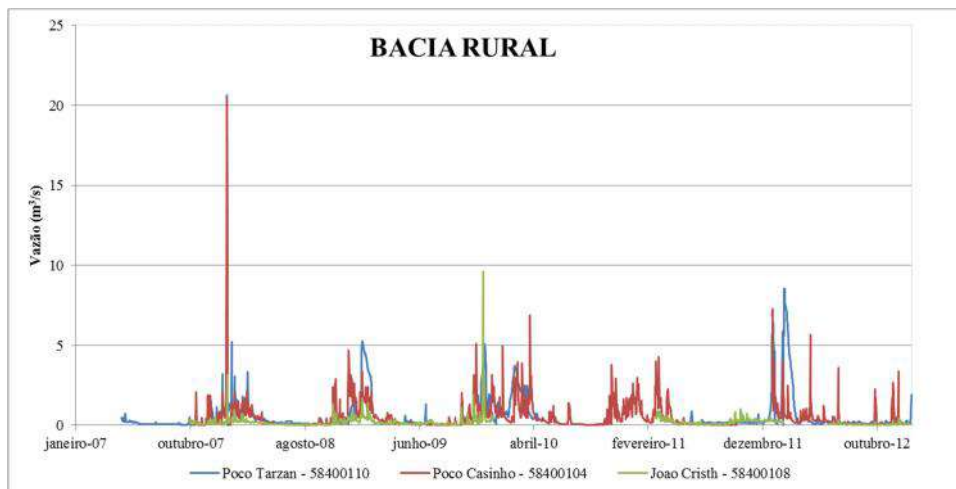


Figura 42 – Vazões diárias para as estações da bacia rural.

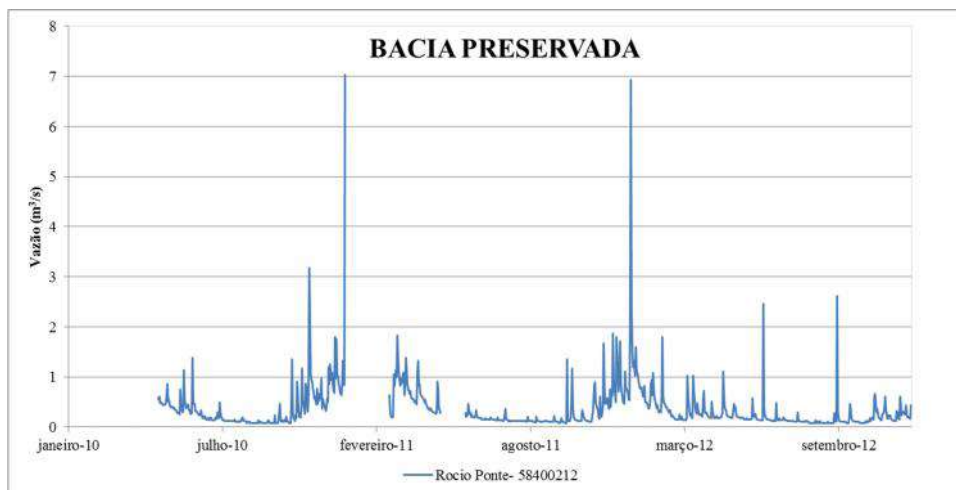


Figura 43 – Vazões diárias para a estação da bacia preservada.

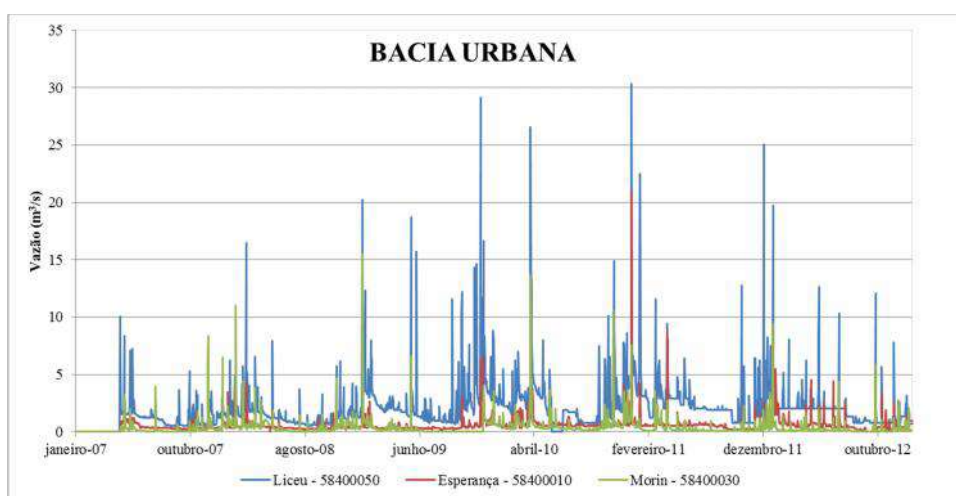


Figura 44 – Vazões diárias para as estações da bacia urbana.

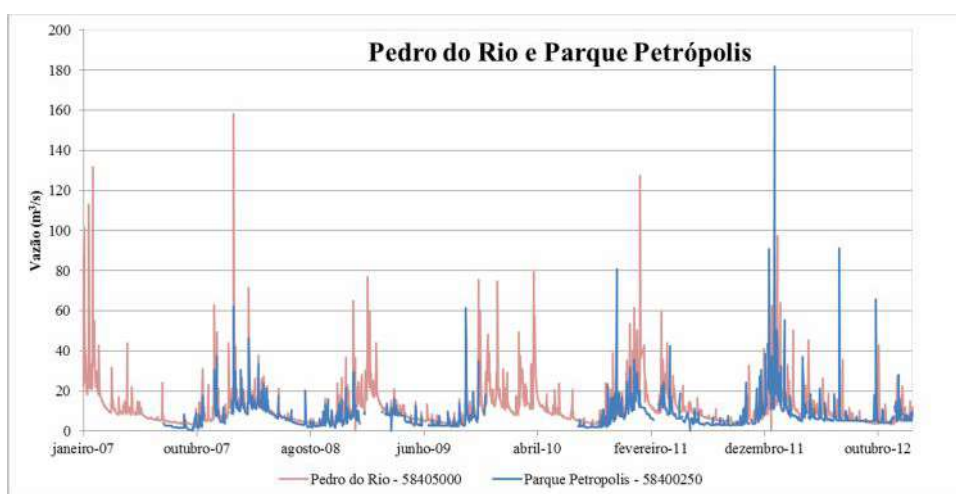
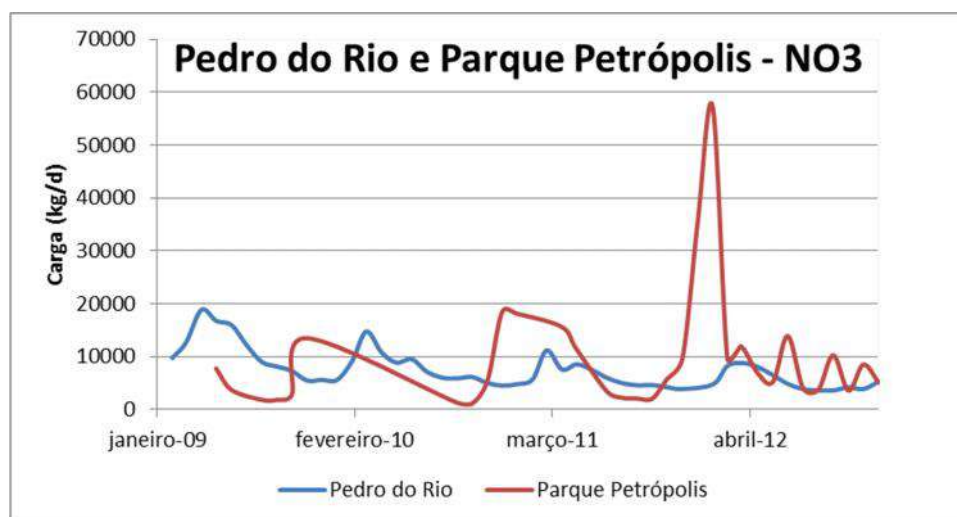
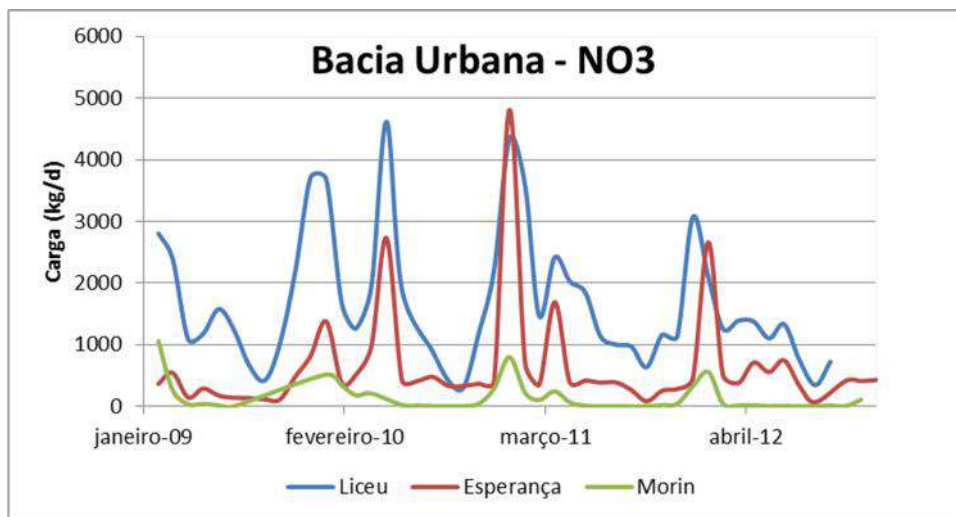


Figura 45 - Vazões diárias para as estações Parque Petrópolis e Pedro Rio.

Com relação aos dados de qualidade de água, para essa aplicação preliminar, foi selecionado o parâmetro nitrato (NO_3) para as 9 estações da RMQAP. Como já apresentado acima, os dados são coletados na bacia com uma frequência que varia de mensal a quadrimestral. Entretanto, para a aplicação, eram necessárias cargas mensais. Dessa forma, com base nas vazões associadas a cada dado de concentração coletado foi possível gerar series de dados mensais usando o programa LOADEST disponibilizado pelo Serviço Geológico americano e comumente usado nas simulações com os modelos SWAT e SPARROW (RUNKEL *et al.*, 2004; TESHAGER *et al.*, 2016; LEE *et al.*, 2016). O programa possui 11 modelos de regressão, calibrados com base em métodos estatísticos específicos, para estimar as cargas dos constituintes em rios usando como base a relação estabelecida entre as vazões e as concentrações dos parâmetros de qualidade de água que são. Não foi possível ajustar um modelo para amônia, assim, ao final apenas foram utilizadas as cargas NO_3 . A seguir são apresentadas as cargas mensais estimadas através do LOADEST para as sete estações selecionadas, organizadas por bacia experimental, para o parâmetro nitrato (Figuras 46, 47, 48 e 49). As estações Parque Petrópolis e Poço Tarzan apresentam os maiores períodos de falhas para o período. Nota-se que a poluição por nitrato, na maior parte do tempo, é mais intensa na estação Parque Petrópolis do que em Pedro do Rio.



Figuras 46 – Cargas mensais das estações Parque Petrópolis e Pedro do Rio para o parâmetro nitrato (NO_3).



Figuras 47 - Cargas mensais das estações Liceu, Morin e Esperança para o parâmetro NO3.



Figura 48 - Cargas mensais da estação Rocio para o parâmetro nitrato NO3.

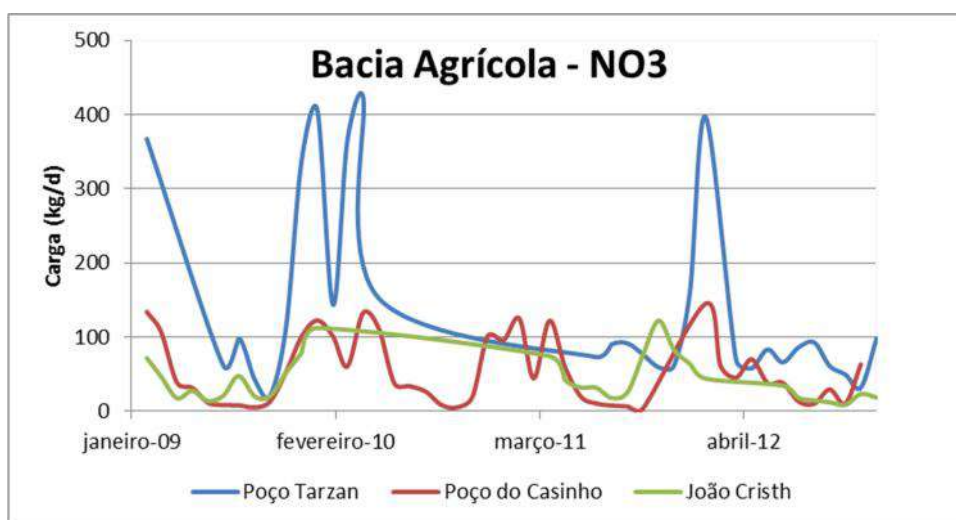


Figura 49 - Cargas mensais das estações Poço Tarzan, João Cristh e Poço do Casinho para o parâmetro nitrato (NO3).

5.2.1.2 Definição de Cenários de Avaliação da RMQA

Nessa etapa da metodologia o objetivo é a definição de cenários para a avaliação do impacto de cada estação da RMQAP no ajuste do modelo. O objetivo dessa avaliação preliminar será redução de custos de operação da rede e a verificação do impacto das bacias experimentais nos dados da rede. Dessa forma, o cenário ótimo, sob o ponto de vista financeiro e de logística, será aquele com o menor número de estações e o cenário mais dispendioso será o que contempla as 9 estações da RMQAP. Assim, para esta primeira aplicação, foram definidos 5 cenários:

- i. Cenário 0 – Cenário padrão – as nove estações da RMQAP

O cenário 0 corresponde ao modelo ajustado em relação aos dados de entrada fornecidos pelo usuário para as nove estações da RMQAP. Em outras palavras, não são realizadas alterações dos parâmetros do modelo ou qualquer procedimento de calibração.

- ii. Cenário IND – Potencial máximo de calibração da estação – as nove estações da RMQAP

O cenário IND refere-se a calibração das estações de forma individual para avaliar o impacto máximo dos dados da estação na representação do modelo e, assim, comparar com o comportamento da estação quando calibrada nos cenários.

- iii. Cenário I – Cenário pessimista – as nove estações da RMQAP

Esse cenário é o cenário atual, com as nove estações de monitoramento, e, objetivamente, o que pretendemos avaliar. Assim, ele será a base de comparação dos demais cenários. Porém, é o que requer a maior quantidade de recursos para ser mantido.

- iv. Cenário II – Cenário otimista – estação Pedro do Rio.

No Cenário I, a RMQAP se reduziria a estação Pedro do Rio (PR). PR fica no exutório da bacia representativa do rio Piabanha e é a estação que reúne as maiores séries de dados. A questão aqui é avaliar a o quão representativa PR é em relação aos

impactos dos usos e ocupação da área de estudo e se ela seria suficiente para simular os dados das demais através dos modelos selecionados.

- v. Cenário III – 5 estações – Pedro do Rio, Parque Petrópolis, Liceu, Poço Tarzan e Rocio

O Cenário III adiciona à configuração do Cenário III as estações que se localizam no exutório de cada bacia experimental: Liceu (L), na bacia urbana, Poço Tarzan (PT, na bacia agrícola e Rocio (R), na bacia preservada. Nesse caso, será avaliada a contribuição da estação do exutório para o ajuste das demais estações de cada bacia experimental e do seu ajuste propriamente dito.

5.2.1.3 Simulação e avaliação dos cenários para vazões e qualidade de água usando o modelo SWAT a partir das métricas estatísticas;

A simulação dos dados de vazão e qualidade de água para área de estudo foi realizada a partir da versão do modelo ArcSWAT 2012.10_3.19. A análise de sensibilidade e calibração do modelo foram executadas com o auxílio do software SWAT-CUP 5.1.6.2. As etapas desenvolvidas compreendem basicamente: a delimitação da bacia, sub-bacias e rede hidrográfica com base no modelo digital de elevação produzido a partir do SRTM e na definição dos exutórios das sub-bacias; definição das Unidades de Respostas Hidrológicas (URH's); alimentação do banco de dados da bacia com os dados climáticos; análise de sensibilidade dos parâmetros do modelo; verificação do modelo definido a partir da compreensão do comportamento hidrológico da bacia simulação; calibração e validação dos cenários.

- Análise de Sensibilidade dos parâmetros

Nesse trabalho optou-se pela análise de sensibilidade global, disponível no SWAT-CUP, e seguiu-se a recomendação de realização de um grande número de simulações para obter resultados consistentes (MORIASI *et al.*, 2007). Assim, com base no manual do modelo, orientações de especialistas e na literatura disponível foi selecionado um conjunto de 13 parâmetros, para as simulações de vazão, para iniciar a análise com suas respectivas faixa de valores, conforme apresentado na Tabela 18. Os

parâmetros encontram-se apresentados seguido pelo seu significado, pelo formato de ajuste SWAT-CUP e da faixa de medição utilizada na análise. Uma descrição detalhada dos parâmetros pode ser encontrada em SWAT (2012) e em ABBASPOUR *et al* (2007). Com auxílio do SWAT-CUP, com o algoritmo SUFI-2, os parâmetros foram testados para a RMQAP toda e para cada estação individualmente.

Tabela 18 – Conjunto de parâmetros selecionados e suas respectivas faixas de medição (ABBASPOUR et al., 2007).

Parâmetros		DESCRIÇÃO DO SIGNIFICADO FÍSICO DO PARÂMETRO	Formato	Faixa de Medição	
				Valor mínimo	Valor máximo
1	v__RCHRG_DP.gw	Fração profunda de percolação do aquífero	absoluto	0	1
2	r__SOL_AWC.sol	Capacidade de água disponível da camada do solo.	relativo	0%	5%
3	v__CH_K2.rte	Condutividade hidráulica efetiva no aluvião do canal principal.	absoluto	25	76
4	v__CH_N2.rte	Coefficiente de Manning para o canal principal.	absoluto	0.025	0.15
5	v__ALPHA_BF.gw	Fator alfa do fluxo base	absoluto	0.1	1
6	r__GW_DELAY.gw	Dias de atraso na água subterrânea	relativo	-20%	20%
7	v__EPCO.hru	Compensação de absorção de plantas	absoluto	0.01	1
8	v__CANMX.hru	Armazenamento máximo das folhas	absoluto	0	15
9	v__ESCO.hru	Fator de compensação da evaporação do solo	absoluto	0.6	0.9
10	v__GW_REVAP.gw	Coefficiente de reabastecimento da água subterrânea	absoluto	0	0.2
11	v__REVAPMN.gw	Profundidade de limiar de água no aquífero raso para o reabastecimento ocorrer	absoluto	0	500
12	v__GWQMN.gw	Profundidade de limiar de água no aquífero raso requerido para que o fluxo de retorno ocorra	absoluto	0	5000
13	r__CN2.mgt	Coefficiente curve number SCS	relativo	-10%	20%

Para a rede foi selecionado um conjunto de 8 parâmetros mais sensíveis apresentados na ordem de sensibilidade: RCHRG_DP, GWQMN, CANMX, ESCO,

CN2, EPCO, GW_REVAP e ALPHA_BF. A análise para cada estação individualmente mostrou que apenas dois parâmetros são comuns ao conjunto de mais sensíveis para todas as estações: RCHRG_DP e GWQMN conforme apresentado na Tabela 19. Esses são dois parâmetros ligados à água subterrânea. Não é possível identificar um conjunto de parâmetros característicos de cada bacia experimental. Pedro do Rio e Parque Petrópolis, as estações mais próximas do exutório da bacia e com o as maiores áreas de drenagem, possuem o mesmo conjunto de parâmetros.

Tabela 19 – Parâmetros mais sensíveis para cada estação da RMQAP.

Pedro do Rio - PR	Parque Petrópolis - PP	Poço Tarzan - PT	Poço do Casinho - PC	João Cristh - JC	Rocio - R	Liceu - L	Esperança - E	Morin - M
v_RCHRG_DP.gw	v_RCHRG_DP.gw	v_RCHRG_DP.gw	v_RCHRG_DP.gw	v_RCHRG_DP.gw	v_RCHRG_DP.gw	v_RCHRG_DP.gw	v_RCHRG_DP.gw	v_RCHRG_DP.gw
v_GWQMN.gw	v_GWQMN.gw	v_GWQMN.gw	v_GWQMN.gw	v_GWQMN.gw	v_GWQMN.gw	v_GWQMN.gw	v_GWQMN.gw	v_GWQMN.gw
v_CANMX.hru	v_CANMX.hru	v_CANMX.hru	r_CN2.mgt	v_ALPHA_BF.gw	v_CANMX.hru	v_ALPHA_BF.gw	v_EPCO.hru	v_CANMX.hru
v_ESCO.hru	v_ESCO.hru	r_CN2.mgt	v_CANMX.hru	r_CN2.mgt	v_ESCO.hru	v_CANMX.hru	v_ESCO.hru	r_CN2.mgt
r_CN2.mgt	r_CN2.mgt	v_ESCO.hru	v_ALPHA_BF.gw	v_CANMX.hru	v_GW_REVAP.gw	v_ESCO.hru	v_GW_REVAP.gw	
v_ALPHA_BF.gw	v_ALPHA_BF.gw				r_CN2.mgt	v_GW_REVAP.gw		
					r_CN2.mgt			

O conjunto de parâmetros mais sensíveis de cada estação foi utilizado na calibração dos cenários. Essa definição é importante pois estudos mostram que o ajuste do modelo está mais relacionado a um bom conjunto de parâmetros do que a algum parâmetro em particular (BEVEN, 2001).

Para a calibração de qualidade de água foram selecionados 10 parâmetros conforme apresentado na Tabela 20. Foram selecionados nove parâmetros para as simulações excluindo-se apenas o BIOMIX.

Tabela 20 – Parâmetros utilizados na análise de sensibilidade de qualidade de água (NO3 e OD).

Parâmetros		DESCRIÇÃO DO SIGNIFICADO FÍSICO DO PARÂMETRO	Formato	Faixa de Medição	
				Valor mínimo	Valor máximo
1	v__NPERCO.bsn	Coeficiente de percolação do nitrogênio	absoluto	0	1
2	v__BIOMIX.mgt	Eficiência de mistura biológica	absoluto	0	1
3	v__CMN.bsn	Fator de taxa para a mineralização da humus de nitrogênio orgânico ativo.	absoluto	0.001	0.003
4	v__SOL_CBN().sol	Teor de carbono orgânico	absoluto	0.05	10
5	v__RCN.bsn	Concentração de nitrogênio na precipitação	absoluto	0	15
6	v__SHALLST_N.gw	Profundidade inicial da água no aquífero raso	absoluto	0	1000
7	v__SOL_NO3(1).	Concentração inicial de NO3 na camada do solo.	absoluto	0	100
8	v__RSDCO.bsn	Coeficiente de decomposição de resíduos	absoluto	0.02	0.1
9	r__SDNCO.bsn	Limite de desnitrificação conteúdo de água	absoluto	-0.7	0.5
10	v__RK2.swq	Taxa de reaeração de oxigênio	absoluto	0	100

- Verificação do Modelo - Compreensão do comportamento Hidrológico da Bacia

A compreensão do comportamento hidrológico da bacia representativa do rio Piabanha através do modelo é fundamental para a definição da sua configuração final a ser utilizada para a simulação de vazões e qualidade da água da área de estudo. Basicamente, podemos separar essa análise em três grupos de acordo com seu objetivo: impacto da delimitação de subbacias, impacto de mapas de uso e mapas de ocupação do solo de diferentes escalas e avaliação das componentes do balanço hídrico através do SWAT Check. As três análises tiveram como base o cenário padrão da área de estudo (detalhado na seção 5.2.1.2) que corresponde aos dados das nove estações da RMQAP resultantes da configuração inicial do modelo sem ajuste dos parâmetros.

i. Impacto da Delimitação das sub-bacias

Nessa análise, foram realizadas diferentes subdivisões da área de estudo em sub-bacias. Para cada subdivisão foi extraído um cenário padrão de vazões que foi comparado com os dados observados das estações da RMQAP com base nas métricas estatísticas definidas para a metodologia na seção 5.2.1.4, quais sejam. Essa análise se justifica pelos diversos estudos acerca do impacto da delimitação das sub-bacias, principalmente, na simulação de vazões. CHO *et* OLIVERA (2009) apresentam uma discussão interessante sobre o tema.

Primeiramente, a área de estudo foi dividida em 9 sub-bacias coincidentes com a área de drenagem de cada estação da RMQA. Esse é o mínimo que po-se ter, já que o objetivo do estudo é calibrar com os dados da Rede para futura avaliação. Nesse caso, não seria possível ter uma única bacia coincidente com a área de estudo como sugerido por MANGUERRA *et* ENGEL *in* CHO *et* OLIVERA (2009). A segunda divisão correspondeu a 45 sub-bacias resultante da combinação entre as sub-bacias sugeridas pelo programa ARC SWAT, as estações da RMQA e a recomendação de sub-bacias com áreas na mesma ordem de grandeza (BRESSIANI, 2016). Foram testadas mais três combinações entre 45 e 9 sub-bacias, chegando-se a divisão final de 31 sub-bacia apresentada na Figura 50 juntamente com o MDE, a hidrografia e a RMQA. A partir dos resultados dessas cinco configurações observou-se que:

- A estação Rocio (R) sofre muito impacto com a subdivisão. O aumento de sub-bacias a montante melhora o ajuste.
- As estações Morin (M) e Poço Tarzan (PT) não são influenciadas;
- As demais sofrem leve impacto com a alteração da configuração da bacia.

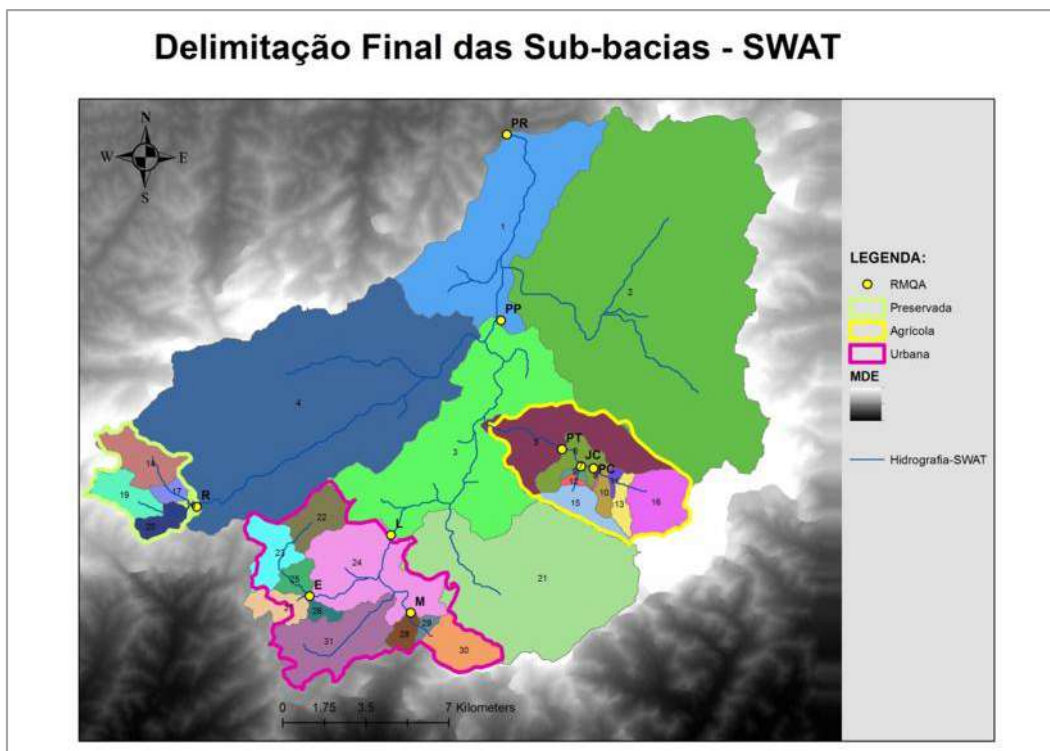


Figura 50 – Mapa da área de estudo com a delimitação final em sub-bacias do SWAT.

ii. Impacto de mapas de uso e ocupação do solo de diferentes escalas

Assim como o impacto da divisão em sub-bacias na simulação de vazões e cargas foi bastante estudado, a resolução dos dados de entrada também gerou diversas pesquisas (CHO *et* OLIVERA, 2009). Dessa forma, com a disponibilização para a bacia de mapas de uso e ocupação do solo em diferentes escalas (1:25.000 e 1:250.000) e o fato da simulação de qualidade de água, a título de investigação, foi realizada a comparação do impacto do detalhamento das classes de uso e ocupação no ajuste do modelo. Similarmente a análise anterior, não foram realizadas alterações nos parâmetros de entrada e foram utilizados os dados de vazão e qualidade de água.

Não foram observados impactos significativos na simulação de vazões das estações analisadas pela melhoria na resolução do mapa de uso e ocupação de solos. Por outro lado, houve algumas diferenças na simulação das cargas de nitrato (NO₃).

Principalmente, nas estações da área agrícola. Vale ressaltar que o mapa de menor escala não apresentava classes referentes a áreas agrícolas. Além disso, o mapa de melhor resolução apresentou um menor percentual de área florestada e um maior percentual de área urbana. De acordo com CHO et OLIVERA (2009), em pequenas bacias, representações espacial mais realistas do dado não necessariamente implicaria numa melhoria do desempenho do modelo na simulação de vazões. Já que o tempo de concentração menor que 24 horas faz com que a bacia inteira drene em um tempo menor que o passo de tempo do SWAT e o modelo capture apenas os efeitos principais no exutório.

iii. Avaliação das componentes do balanço hídrico através do SWAT Check

Por último, foi realizada uma verificação das diversas componentes do balanço hídrico no programa SWAT Check que apresenta as médias anuais obtidas na simulação com o SWAT. Da mesma forma, em princípio, não foram alterados os parâmetros de entrada e, foram utilizadas as nove estações da RMQA. A Figura 51 apresenta os resultados da simulação do SWAT no SWAT Check.

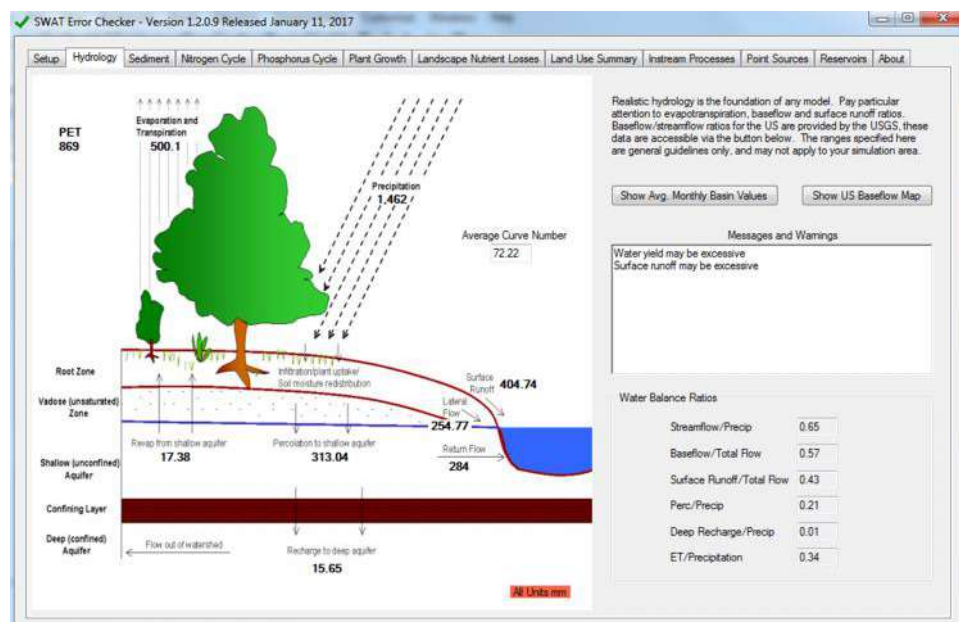


Figura 51 – Resultado da simulação do SWAT no SWAT Check sem alteração de parâmetros para o método Penman-Monteith.

Observam-se alertas sobre escoamento superficial e sub-superficial excessivos e a evapotranspiração um pouco abaixo dos valores apresentados na literatura. Não foram

encontrados estudos com valores detalhados para a bacia do Piabanha, apenas estimativas dos valores médios de evapotranspiração real e potencial, em torno de 700mm/ano e 950mm/ano respectivamente (GONÇALVES, 2008, BARROS, 2012, ARAÚJO, 2016). No SWAT, a evapotranspiração potencial pode ser calculada por três métodos diferentes: Penman-Monteith (MONTEITH, 1965; ALLEN *et al.*, 1989), o método padrão, Priesley-Taylor (PRIESLEY *et* TAYLOR, 1972) e Hargreaves (HARGREAVES, G.H. *et* SAMANI, 1985) ou pode ser inserida pelo usuário. Assim, foram realizados testes com os outros dois métodos, Priesley-Taylor (PRIESLEY *et* TAYLOR, 1972) e Hargreaves, cujos resultados no SWAT Check encontram-se apresentados respectivamente nas Figuras 52 e 53. Observa-se que ambos os métodos elevam o valor da evapotranspiração. Com relação a evapotranspiração real o método Hargreaves apresenta um valor mais próximo da estimativa encontrada para a bacia (700mm/ano). Entretanto, ele superestima a evapotranspiração potencial enquanto que o Priesley-Taylor apresenta um valor compatível com o estimado (950mm/ano). BARROS (2012) comparou esses dois métodos de estimativa de evapotranspiração potencial para a bacia do Piabanha e concluiu que o Priesley-Taylor apresentou melhores resultados. Além disso, durante o processo de calibração, com o ajuste dos parâmetros é possível elevar o valor da evapotranspiração real. Dessa forma, optou-se em utilizar o método Priesley-Taylor para a aplicação da metodologia.

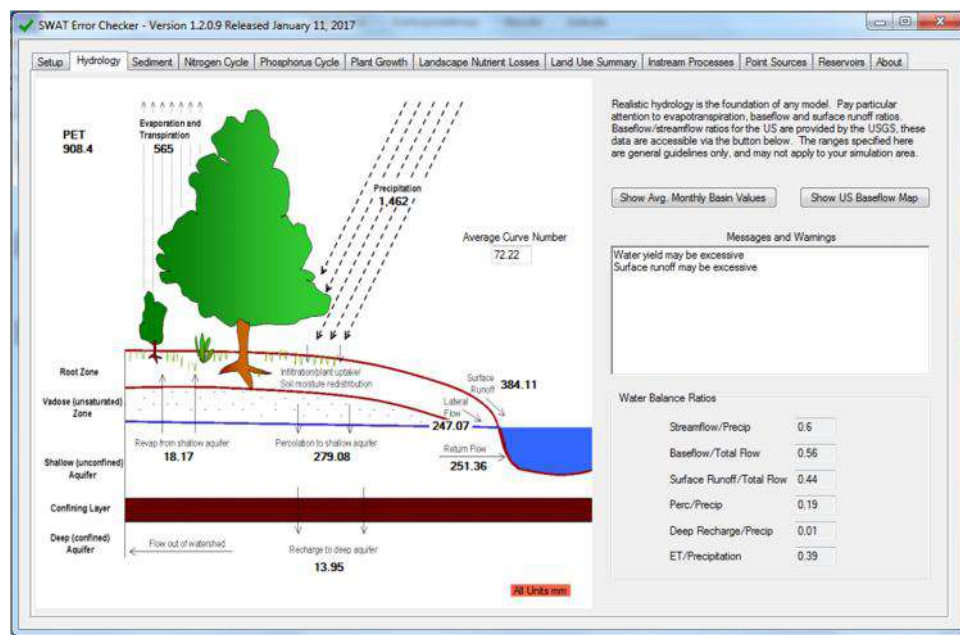


Figura 52 - Resultado da simulação do SWAT no SWAT Check sem alteração de parâmetros para o método Priesley-Taylor.

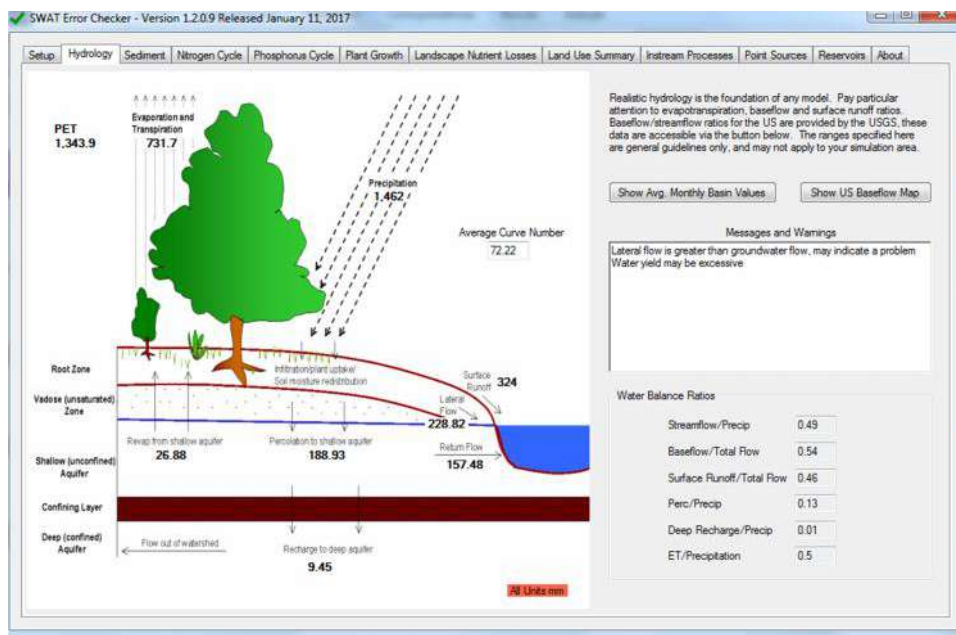


Figura 53 - Resultado da simulação do SWAT no SWAT Check sem alteração de parâmetros para o método Hargreaves.

Vale ressaltar que o excesso de escoamento superficial e sub-superficial será corrigido por meio do ajuste dos parâmetros durante o processo de calibração, bem como, outras possíveis incompatibilidades físicas e conceituais. Dessa maneira, o programa SWAT Check funciona como ferramenta de apoio até o ajuste final do modelo.

- Calibração e Validação dos cenários

Nesse trabalho, foi utilizada uma calibração semi-automática com auxílio do programa SWATCup (algoritmo SUFI2), ARC SWAT e SWAT Check para os cinco quatro definidos na seção 5.2.1.2. Foram utilizados os parâmetros definidos para cada cenário na análise de sensibilidade anteriormente apresentada. A calibração foi realizadas na escala mensal para os dados de vazão e para o parâmetros de qualidade de água nitrato. Não foi realizada validação temporal em razão da serie ser pequena e, assim, ser recomendável a utilização de todos os dados para a calibração (ARNOLD *et al.*, 2012). Além disso, a metodologia proposta consiste em uma espécie de validação espacial.

Foi realizada uma tentativa de calibração diária para os dados de vazão, entretanto, o ajuste foi insatisfatório para todas estações com exceção de Pedro do Rio e

Parque Petrópolis. A escala diária seria mais apropriada a aplicação da metodologia já que a área estudo é pequena. . Foram definidos dois anos de período de inicialização do modelo para a calibração (2007 e 2008) para permitir a minimização de possíveis efeitos das estimativas iniciais das variáveis (SANTHI *et al.*, 2008)

A Bacia Representativa do rio Piabanha foi dividida em 31 sub-bacias (Figura 50), conforme apresentado na seção 5.2.2.1, que a partir da combinação dos dados de uso e ocupação do solo, solos e declividade, resultaram em 370 unidades de resposta hidrológica (HRU).

Para o cenário I foram realizadas dois tipos de calibração: múltipla e única de forma a comparar os resultados. No primeiro caso, é realizada a calibração simultânea de todas as estações no SWAT Cup. O programa permite realizar o ajuste dos parâmetros com base nos dados observados de múltiplas estações. Na calibração única, cada estação é calibrada separadamente. As estações são calibradas de montante para jusante. Primeiro, são calibradas as estações chamadas “independentes” ou de cabeceira que são aquelas que não possuem nenhuma outra estação a montante. Em seguida, o modelo é ajustado para essas estações utilizando os valores dos parâmetros calibrados apenas para as bacias de contribuição das estações independentes. A seguir, a área de drenagem entre a estação independente e a próxima estação a jusante, cuja área inclui a área da estação independente, é calibrada. E assim por diante até o exutório da bacia.

Para ilustrar esse procedimento observe o diagama unifilar da RMQAP representado na Figura 54. As estações em vermelho correspondem as independentes, e são calibradas primeiramente. Em seguida, as estações verdes são calibradas. A seguir a amarela e por ultimo a azul. O segundo método tende a resultar num melhor ajuste para as estações independentes, mas por outro lado é mais dispendioso. Como a área de estudo corresponde a uma bacia pequena a diferença entre os dois métodos pode ser pouco significativa. Pois o cenário ideal para a calibração múltipla é quando são usadas apenas estacoes independentes para a calibração. Quando há estações dependentes corre-se o risco de deixar de calibrar uma estação em função da outra (BOWDEN *et al.*, 2006).

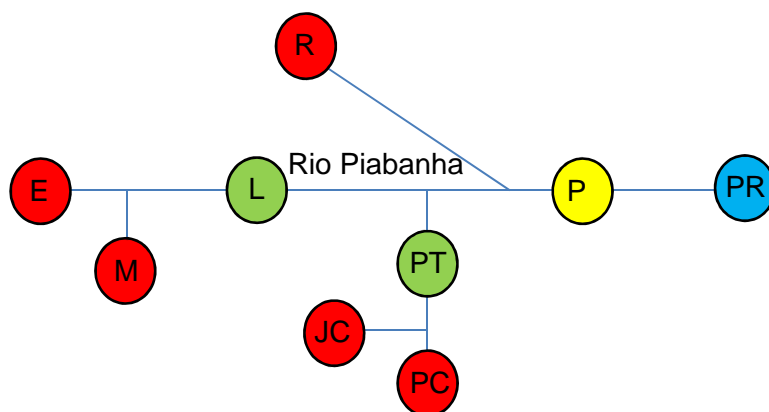


Figura 54 - Diagrama unifilar da RMQAP e etapas de calibração do método de calibração única por cores.

A calibração do parâmetro de qualidade de água, nitrato, foi realizada a partir dos cenários calibrados para vazão de forma múltipla e para cada estação individualmente.

Não foi realizada validação temporal, pois a série de dados era curta. Não foi realizada calibração de sedimentos por não estarem disponíveis dados para a bacia.

Durante a calibração de vazões foram incluídos mais dos parâmetros, além dos selecionados na análise de sensibilidade, pois o modelo estava anulando o escoamento de base da bacia e transferindo grande parte para o sub-superficial, superestimando esse último. É sabido que o escoamento subsuperficial ocupa papel significativo na bacia, mas não foram encontrados estimativas de valores para ele (ARAÚJO, 2016). Isso pode ter ocorrido em função da deficiência do detalhamento do mapa de solos da bacia. Dessa forma com base na literatura, foram incluídos, manualmente, os parâmetros SLSUBBSN e HRU_SLP relacionados à inclinação média e o comprimento médio da inclinação do escoamento sub-superficial e (GITAU *et* CHAUBEY, 2010; QIU *et al.*, 2012). Eles foram calibrados de forma relativa e suas faixas de valores foram, respectivamente, +40-50% e -40%-50%. Além disso, fixou-se o limite superior do RCHRG_DP em 0.7 para a toda a bacia com exceção bacia urbana.

- Calibração dos cenários

As métricas estatísticas selecionadas para esse estudo, que passam a ser descritas a seguir, são as comumente utilizadas nas aplicações do modelo SWAT (MORIASI *et al.*, 2007; BRESSIANI, 2016). Elas inclusive foram usadas em alguns dos estudos para

avaliação de RMQA com base nos modelos SWAT e SPARROW (DE GIROLAMO *et al.*, 2003;; PURI *et al.*, 2017) e estão entre as recomendadas por ASABE (2017).

A primeira delas, uma das mais populares dentre as aplicações com o modelo SWAT, é o coeficiente de eficiência Nash e Sutcliffe (NSE) calculado a partir da Eq. 13. A segunda o PBIAS que mede o percentual da tendência média dos dados simulados serem maiores ou menores que os dados observados (MORIASI *et al.*, 2007). Valores positivos significam subestimativa do modelo e negativos a superestimativa. Quanto mais próximo do zero melhor o ajuste do modelo. Ele pode ser calculado através da Eq. 14.

$$PBIAS = 1 - \frac{\sum_{i=1}^n Y_i^{obs} - Y_i^{sim} * 100}{\sum_{i=1}^n Y_i^{obs}} \quad (14)$$

As duas últimas métricas são o Coeficiente de Determinação (R^2) e a razão entre a raiz do erro médio quadrático (RMSE) pelo desvio padrão das observações conhecida como RSR. O R^2 é muito utilizado para avaliação de modelos, entretanto, é bastante sensível a *outliers*. Quanto mais próximo de um melhor o ajuste do modelo. Já o RSR avalia a variação do resíduo da simulação. Então, quanto mais próximo de zero melhor, significando que não há resíduo.

A avaliação do desempenho do modelo SWAT será feita com base nos valores estabelecidos na literatura para a avaliação do desempenho do modelo com base nas métricas estatísticas, para vazão (Q) e nutrientes (NO3) apresentados na Tabela 21 (MORIASI *et al.*, 2015). As tabelas com os resultados serão apresentadas pintadas conforme a avaliação do desempenho do modelo em relação à métrica estatística: verde para o um desempenho “muito bom”, amarelo para “bom”, laranja para “satisfatório e branco para “insatisfatório”. Vale ressaltar, que esses valores são apenas orientações e não estabelecem padrões estritos de rejeição e aceitação do modelo (ASABE, 2017). Para a avaliação da RMQAP eles serão usados de forma mais relativa do que absoluta como apresentado no seção 5.1.

Tabela 21 - Valores das métricas estatísticas para avaliação do desempenho do modelo.

Avaliação do desempenho	R2	RSR	NS (Q)	NS (NO3)	PBIAS(%) (Q)	PBIAS(%) (NO3)
Muito bom	-	0,00 RSR 0,50	0,80<NSE 1,00	0,65<NSE 1,00	PBIAS<±5	PBIAS<±15
Bom	-	0,50<RSR 0,60	0,70<NSE 0,80	0,50<NSE 0,65	±5 PBIAS<±10	±15 PBIAS<±20
Satisfatório	RSR>0,70	0,60<RSR 0,70	0,50<NSE 0,70	0,35<NSE 0,50	±10 PBIAS<±15	±20 PBIAS<3015
Insatisfatório	RSR<0,70	RSR>0,70	NSE 0,50	NSE 0,35	PBIAS ±15	PBIAS ±30

- Resultados da modelagem hidrológica

Os resultados obtidos na aplicação da metodologia proposta com base na modelagem hidrológica usando o SWAT serão apresentados, primeiramente, para a calibração das vazões e, em seguida, para calibração do parâmetro de qualidade de água nitrato (NO₃) .

Nas tabelas a seguir são apresentadas as métricas estatísticas de avaliação da RMQAP. “IND” refere-se à calibração das estações individualmente (seção 5.2.1). O cenário “0” é o cenário padrão, isto é, a saída do modelo não calibrado. O cenário “I-REDE” refere-se ao cenário I calibrado de forma múltipla, ou seja, com todas as estações sendo calibradas simultaneamente (seção 5.2.2.1). O cenário “I-IND” refere-se ao cenário I calibrado de forma única (seção 5.2.2.1). O cenário II foi calibrado apenas para a estação PR. E o modelo calibrado foi aplicado às demais estações da rede sem alteração dos parâmetros numa espécie de validação espacial. O cenário IV foi calibrado para as estações L, R, PT, PR e PP e o modelo ajustado foi aplicado às estações E, M, PC e JC, sem ajuste de parâmetro como no caso do cenário II.

As Tabelas 22, 23, 24 e 25 apresentam os valores de NS, PBIAS, R² e RMS, respectivamente, para cada um dos cenários definidos e para cada estação de monitoramento da RMQAP referente à calibração de vazões com base no modelo SWAT.

Tabela 22 - Valores de NS para a calibração de vazões para os cenários definidos.

NASH SUTCLIFFE	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	III
Pedro do Rio - PR	0.77	0.34	0.75	0.74	0.77	0.74
Parque Petrópolis - PP	0.67	0.47	0.63	0.67	0.63	0.67
Poço Tarzan - PT	0.65	0.64	0.60	0.65	0.60	0.65
Poço do Casinho - PC	0.73	0.73	0.58	0.73	0.57	0.67
João Cristh - JC	0.51	0.49	0.46	0.51	0.44	0.48
Rocio- R	0.71	0.02	0.57	0.71	0.56	0.71
Liceu - L	0.48	0.39	0.32	0.42	0.32	0.48
Esperança - E	0.54	0.35	0.44	0.54	0.44	0.55
Morin - M	0.43	0.36	0.23	0.43	0.20	0.31

Tabela 23 - Valores de PBIAS para a calibração de vazões para os cenários definidos.

PBIAS	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	III
Pedro do Rio - PR	7	-1	6	13	7	14
Parque Petrópolis - PP	16	13	18	22	19	22
Poço Tarzan - PT	3	-5	5	3	5	3
Poço do Casinho - PC	33	33	37	33	36	37
João Cristh - JC	21	23	27	21	28	27
Rocio- R	-2	-38	-26	-2	-26	-2
Liceu - L	16	26	32	16	32	16
Esperança - E	17	17	24	17	24	19
Morin - M	25	32	36	25	37	33

Tabela 24 - Valores de R^2 para a calibração de vazões para os cenários definidos.

R2	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	III
Pedro do Rio - PR	0.85	0.85	0.85	0.84	0.85	0.84
Parque Petrópolis - PP	0.75	0.74	0.73	0.79	0.73	0.79
Poço Tarzan - PT	0.66	0.64	0.65	0.66	0.66	0.66
Poço do Casinho - PC	0.92	0.92	0.91	0.92	0.91	0.92
João Cristh - JC	0.56	0.56	0.55	0.56	0.55	0.56
Rocio- R	0.75	0.72	0.73	0.75	0.72	0.75
Liceu - L	0.77	0.75	0.77	0.77	0.77	0.76
Esperança - E	0.69	0.64	0.67	0.69	0.67	0.69
Morin - M	0.60	0.54	0.53	0.60	0.51	0.57

Tabela 25 - Valores de RSR para a calibração de vazões para os cenários definidos.

RSR	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	III
Pedro do Rio - PR	0.48	0.81	0.50	0.51	0.48	0.51
Parque Petrópolis - PP	0.58	0.73	0.61	0.58	0.61	0.58
Poço Tarzan - PT	0.59	0.60	0.63	0.59	0.63	0.59
Poço do Casinho - PC	0.52	0.52	0.65	0.52	0.65	0.58
João Cristh - JC	0.70	0.71	0.74	0.70	0.75	0.72
Rocio- R	0.54	0.99	0.66	0.54	0.66	0.54
Liceu - L	0.72	0.78	0.83	0.76	0.83	0.72
Esperança - E	0.68	0.81	0.75	0.68	0.75	0.67
Morin - M	0.75	0.80	0.88	0.75	0.90	0.83

Com relação à calibração de vazões é possível observar que a maioria das estações, individualmente, para os parâmetros NS, R^2 e RMS, teve desempenho acima de satisfatório. O parâmetro PBIAS teve resultados absolutos menos favoráveis.

Entretanto, vale ressaltar que o critério utilizado nesse estudo é recente e mais restritivo para os valores de PBIAS em relação ao critério antigo usualmente utilizado nas aplicações do SWAT e indicado por MORIASI *et al.* (2007) .

As estações da bacia urbana de um modo geral apresentaram ajuste insatisfatório para a maior parte dos cenários para as todas as métricas estatísticas. A estação Esperança foi a com melhor desempenho. Por outro lado, a bacia agrícola foi a com melhor desempenho.

Nota-se, a partir do cenário 0 (padrão), que a calibração não teve impacto considerável na simulação das vazões das estações da bacia agrícola. As estações PC e PT já apresentam bons ajustes nesse cenário. Por outro lado, as estações R e PR sofreram impacto significativo. A estação R, por exemplo, apresentou o maior impacto. Ela melhorou todas as métricas estatísticas, que demonstravam desempenho “insatisfatório” no cenário 0, com exceção de R^2 , que já era satisfatório.

Comparando os resultados dos cenários I-REDE e I-IND, sob o ponto de vista do tipo de calibração, é possível notar diferenças substanciais na estação R, PC e E para todas as métricas com exceção de R^2 . As três estações são ditas independentes e acabaram sendo despriorizadas, no cenário I-REDE, em função das demais, mesmo em uma pequena bacia, em consonância com os resultados de BOWDEN *et al.* (2006). Talvez a calibração múltipla seja mais adequada para bacias com uso e cobertura do solo homogêneas. Os resultados do cenário I –REDE e II são bem similares parecendo que a calibração múltipla prioriza a estação PR que possui a maior área de contribuição.

Com relação ao cenário II, em que apenas a estação PR é utilizada para a calibração do modelo SWAT, os resultados foram um pouco menos favoráveis. Esse cenário representa uma forma comumente observada em aplicações do modelo SWAT, que utiliza apenas a estação do exutório da bacia para calibração da área toda (ZHANG *et al.*, 2008; NIRAULA *et al.*, 2012). A partir desse cenário é possível observar que as demais estações sofrem um significativo decréscimo em todas as suas métricas em comparação a calibração individualizada (cenário I-IND), com destaque para R e PC. Esses resultados estão de acordo com os encontrados por NIRAULA *et al.* (2012) e demonstram, que mesmo em uma bacia com pequenas dimensões, é importante a inclusão de outras estações além da do exutório para uma melhor representação do modelo.

No cenário III pretende-se avaliar como as estações do exutório das bacias experimentais estão se relacionando as estações das suas respectivas áreas de contribuição e as estações externas à bacia. Observa-se um decréscimo no desempenho da estação M, já a estação E é bem representada pela estação L. N bacia agrícola as estações PC e JC são bem representadas por PT.

As Tabelas 26, 27, 28 e 29 apresentam os valores de NS, PBIAS, R2 e RSR, respectivamente, para cada um dos cenários definidos e para cada estação de monitoramento da RMQAP referente à calibração de nitrato (NO3) com base no modelo SWAT.

A calibração para o parâmetro de qualidade de água nitrato (NO3) de um modo geral não apresentou resultados satisfatórios para todas as métricas R². As estações JC foi a que apresentou melhores resultados.

Tabela 26 - Valores de NSE para a calibração de NO3 para os cenários definidos.

NASH SUTCLIFFE	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	IV
Pedro do Rio - PR	-1.77	-26.46	-2.95	-3.29	-1.77	-2.92
Parque Petrópolis - PP	0.17	-0.32	-0.56	-0.61	-0.42	-0.54
Poço Tarzan - PT	0.22	-65.92	0.11	0.18	0.06	0.22
Poço do Casinho - PC	0.70	-262.18	0.52	0.70	-23.56	-3.07
João Cristh - JC	-0.42	-230.31	-0.68	-0.42	-0.87	-1.37
Rocio- R	0.42	-0.54	0.31	0.42	0.16	0.42
Liceu - L	-0.12	-4.49	-1.47	-1.43	-1.35	-0.12
Esperança - E	0.26	-0.82	-1.07	0.26	-0.61	-0.92
Morin - M	0.55	-1.74	-0.23	0.55	-0.30	-0.21

Tabela 27 - Valores de PBIAS para a calibração de NO3 para os cenários definidos.

PBIAS	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	IV
Pedro do Rio - PR	50	-60	81	88	50	79
Parque Petrópolis - PP	42	45	91	95	83	92
Poço Tarzan - PT	53	-315	64	58	43	53
Poço do Casinho - PC	28	-586	26	15	-128	-36
João Cristh - JC	66	-319	71	66	55	56
Rocio- R	29	57	41	29	34	29
Liceu - L	31	-2	75	84	49	31
Esperança - E	22	22	79	22	59	73
Morin - M	7	-34	85	7	57	85

Tabela 28 - Valores de R^2 para a calibração de NO_3 para os cenários definidos.

R2	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	IV
Pedro do Rio - PR	0.10	0.20	0.10	0.07	0.10	0.09
Parque Petrópolis - PP	0.31	0.10	0.05	0.20	0.09	0.12
Poço Tarzan - PT	0.46	0.22	0.59	0.57	0.26	0.46
Poço do Casinho - PC	0.86	0.09	0.60	0.68	0.07	0.20
João Cristh - JC	0.28	0.03	0.16	0.28	0.07	0.08
Rocio- R	0.56	0.16	0.54	0.56	0.48	0.56
Liceu - L	0.38	0.11	0.05	0.11	0.10	0.38
Esperança - E	0.39	0.23	0.08	0.39	0.16	0.11
Morin - M	0.56	0.01	0.04	0.56	0.00	0.13

Tabela 29 - Valores de RSR para a calibração de NO_3 para os cenários definidos.

RSR	IND.	CENÁRIOS				
ESTAÇÃO		0	I - REDE	I-IND.	II	IV
Pedro do Rio - PR	1.66	5.24	1.99	2.07	1.66	1.98
Parque Petrópolis - PP	0.91	1.15	1.25	1.27	1.19	1.24
Poço Tarzan - PT	0.88	-2.99	0.94	0.91	0.97	0.88
Poço do Casinho - PC	0.55	0.15	0.69	0.59	4.96	2.02
João Cristh - JC	1.19	0.24	1.30	1.19	1.37	1.54
Rocio- R	0.76	1.24	0.83	0.76	0.92	0.76
Liceu - L	1.06	2.34	1.57	1.56	1.53	1.06
Esperança - E	0.86	1.35	1.44	0.86	1.27	1.39
Morin - M	0.67	1.66	1.11	0.67	1.14	1.10

5.2.2 Simulação dos dados de vazão para previsão dos dados de qualidade de água usando redes neurais;

A previsão dos dados do parâmetro de qualidade nitrato através do ajuste de uma rede neural foi realizado, através do software MATLAB R2012a, tendo como base o procedimento recomendado por BERRY *et* LINOFF (2004) apresentado da seção 2.5.1, e as etapas de 1 a 6:

1. Identificação dos dados de entrada e saída;
2. Pré-processamento com a transformação dos dados no intervalo entre +1 e -1;
3. Configuração da topologia da rede;

4. Treinamento da rede em um conjunto representativo de exemplos de treinamento.

5. Separação dos dados em um conjunto de teste e outro de validação e usar o conjunto de validação para definir os pesos que minimizam o erro;

6. Avaliação da rede usando o conjunto de teste;

Os dados de entrada consistiram nos dados de vazão média diária, de cada estação, referente ao dia que foi realizada a coleta que gerou o dado de nitrato. Os dados de saída para calibração da rede referem-se aos dados de nitrato referentes a cada uma das nove estações da rede referentes às coletas no período de 2009 a 2012 cujas estatísticas encontram-se apresentados na Tabela 5.

Os dados de entrada e saída foram pré-processados através da padronização dos valores no intervalo de -1 a 1. Não foram retirados dados que poderiam ser *outliers* pelas mesmas razões já apresentadas na seção 4.2.1.

A definição da topologia da rede foi feita de modo a ajustar uma rede mais o simples para facilitar o uso da metodologia. Foi definida uma rede neural *feedforward* com apenas uma camada oculta, com dez neurônios, com a função tangente-sigmoide nas camadas oculta e uma função linear na camada de saída. Os valores iniciais de pesos e bias iniciais foram determinados usando o método Nguyen-Widrow (NGUYEN E WIDROW, 1990). Os dados foram divididos nos conjuntos teste, validação e treinamento de forma randômica.

A Tabela 30 apresenta as métricas estatísticas usadas para avaliar o desempenho da rede neural que correspondem às mesmas utilizadas na modelagem hidrológica definidas na seção 5.2.1.3 assim como os mesmos critérios de avaliação e cores. Os resultados para as todas as estações em todas as métricas, com exceção de Liceu e o PBIAS de Esperança, foram minimamente satisfatórios.

Tabela 30 – Avaliação do desempenho das redes neurais com base nas métricas estatísticas

REDES NEURAIS				
ESTAÇÃO	NSE	PBIAS	R2	RSR
Pedro do Rio - PR	0.65	23	0.66	0.59
Parque Petrópolis - PP	0.73	20	0.73	0.52
Poço Tarzan - PT	0.52	26	0.53	0.69
Poço do Casinho - PC	0.98	12	0.98	0.13
João Cristh - JC	0.72	15	0.73	0.53
Rocio- R	0.74	10	0.75	0.51
Liceu - L	0.22	64	0.44	0.88
Esperança - E	0.59	42	0.62	0.64
Morin - M	0.73	28	0.78	0.52

5.2.3 Estimativa do impacto de cada estação a partir do calculo do

$$I_{RMQAP}^e$$

A estimativa do impacto da cada estação da RMQAP foi realizada a partir do calculo do I_{RMQAP}^e definido pela Eq. 12. As Tabelas 31 e 32 apresentam os erros relativos (ER_Q e ER_{QA}) padronizados no intervalo de 0 a 1 para a calibração do modelo SWAT de vazões e de nitrato (NO_3) respectivamente tendo como base o NSE do cenário IND conforme detalhado na seção 5.1 (Tabela 30). Os resultados para o I_{RMQAP}^e encontram-se apresentados na Tabela 33 ordenados de forma decrescente acompanhados das áreas incrementais das estações (AI) e dos coeficientes de ponderação F1 e F2 calculados com base na media do NSE e R^2 do ajuste das redes neurais. Observa-se que Pedro do Rio (PR) é a estação com maior impacto na calibração do SWAT e Poço Tarzan (PT) a menos relevante.

Tabela 31 – Erros relativos padronizados entre 0 e 1 para a calibração de vazão usando o modelo SWAT.

ERRO RELATIVO		CENÁRIOS - VAZÃO				
ESTAÇÃO	IND.	0	I - REDE	I-IND.	II	III
Pedro do Rio - PR	0.77	0.57	0.00	0.31	0.00	0.14
Parque Petrópolis - PP	0.67	0.31	0.08	0.00	0.11	0.00
Poço Tarzan - PT	0.65	0.02	0.12	0.00	0.14	0.00
Poço do Casinho - PC	0.73	0.00	0.41	0.00	0.41	0.29
João Cristh - JC	0.51	0.04	0.16	0.00	0.26	0.21
Rocio- R	0.71	1.00	0.39	0.00	0.39	0.00
Liceu - L	0.48	0.19	0.70	1.00	0.62	0.00
Esperança - E	0.54	0.36	0.36	0.00	0.35	0.07
Morin - M	0.43	0.17	1.00	0.00	1.00	1.00

Tabela 32 - Erros relativos padronizados entre 0 e 1 para a calibração de nitrato usando o modelo SWAT.

ERRO RELATIVO	IND.	CENÁRIOS - NO3				
ESTAÇÃO		0	I - REDE	I-IND.	II	III
Pedro do Rio - PR	-1.77	0.02	0.04	0.08	0.00	0.12
Parque Petrópolis - PP	0.67	0.00	0.37	0.42	0.10	0.78
Poço Tarzan - PT	0.50	0.55	0.02	0.02	0.02	0.00
Poço do Casinho - PC	0.59	0.68	0.00	0.00	1.00	1.00
João Cristh - JC	0.83	1.00	0.03	0.00	0.03	0.42
Rocio- R	0.59	0.00	0.00	0.00	0.02	0.00
Liceu - L	0.00	0.06	1.00	1.00	0.30	0.00
Esperança - E	0.54	0.00	0.44	0.00	0.10	0.84
Morin - M	0.00	0.00	0.11	0.00	0.04	0.26

Tabela 33 - Impacto das estações da RMQAP estimado pelo I_{RMQAP} .

ESTAÇÃO	IRMQAP	AI (%)	F1	F2
Pedro do Rio - PR	18.8	36.3	0.7	0.3
Parque Petrópolis - PP	10.9	44.8	0.7	0.3
Rocio- R	3.0	3.1	0.7	0.3
Liceu - L	2.6	6.5	0.3	0.7
Poço do Casinho - PC	1.9	2.3	1.0	0.0
Morin - M	1.8	1.7	0.8	0.2
Esperança - E	1.4	3.5	0.6	0.4
João Cristh - JC	0.3	1.0	0.7	0.3
Poço Tarzan - PT	0.1	0.8	0.5	0.5

5.2.4 Discussão dos resultados e avaliação da RMQAP.

A avaliação da RMQAP será feita com base nos resultados obtidos na calibração do modelo SWAT para vazões e qualidade de água, nas redes neurais e na estimativa do I_{RMQAP} .

Com relação à calibração das vazões no SWAT observa-se que as estações Pedro do Rio (PR) e Parque Petrópolis (PP) oferecem grande impacto na calibração do modelo em todos os cenários. O cenário II mostra que PR representa bem a estação PP, ambas localizadas no curso principal da bacia. Já na calibração do nitrato, com base nos erros relativos apresentados na Tabela 32, é possível observar que a calibração de PR apresentou pouco impacto que pouco variou ao longo dos cenários. Ao contrário de PP que sofre prejuízos consideráveis nos cenários I e III. Inclusive PR apresentou NSE negativo no cenário IND zerando a parcela do impacto advindo da calibração de

qualidade de água no cálculo do I_{RMQAP} . As estações são a que apresentam as maiores áreas incrementais da bacia, muito maiores em relações as demais e, assim, as mais impactantes. O que proporcionou uma elevação nos valores de impacto estimados através do I_{RMQAP} . As duas estações devem ser mantidas na RMQAP. Apesar de o cenário II ter demonstrado que Pedro do Rio (PR) não consegue representar bem a bacia como um todo sendo necessárias estações a maior parte das estações das bacia experimentais. Ou seja, ela não é representativa do impacto dos diferentes usos do solo na bacia.

Os dados das estações da bacia agrícola oferecem pouco impacto na calibração de vazões, com exceção de Poço do Casinho nos cenários I-REDE e II. Por outro lado, ela tem seu ajuste pouco afetado no cenário III demonstrando que PT consegue representá-la bem. As estações PC e PT já apresentam um ajuste satisfatório no cenário 0. A estação PT, mesmo na área agrícola, pode ser representada com desempenho satisfatório em qualquer cenário (com exceção da métrica R^2). Mesmo no cenário 0 o modelo a representa satisfatoriamente. Já o Cenário II mostra que as estações Poço Tarzan (PT) e João Cristh (JC), assim como no cenário I-REDE, tem seu ajuste pouco prejudicado demonstrando que podem ser representadas pela estação PR.

Com relação à calibração do nitrato nota-se que a estação PC apresenta uma ajuste “Bom” no cenário IND representando, assim, um grande impacto em relação ao cenário 0. E ela continua sofrendo prejuízo nos demais cenários. Mas no caso do nitrato, PT não oferece uma boa representação da estação. Confirmando, assim, que seu comportamento, tanto para vazões quanto nitrato, é bem particular e difícil de ser representado por outra estação da rede (vide erros relativos nas Tabelas 31 e 32. A estação PT também possui comportamento similar ao que ocorre na calibração de vazões: não sofre prejuízo significativo na sua calibração nos cenários I, II, II (vide erros relativos nas Tabelas 31 e 32) podendo assim ser representada pelas estações a jusante. A estação JC possui um comportamento variável, sofre pouco prejuízo no cenário II e um prejuízo considerável no cenário IV. Aqui, também, não consegue representar bem a estação. Entretanto, como o NSE de JC deu negativo seu impacto na calibração de qualidade de água não será considerado no cálculo do I_{RMQAP} . Com relação ao índice, ele demonstrou que a estação mais importante para a bacia agrícola é a estação PC e a menos relevante é PT quase com o mesmo valor de impacto de JC que vem em seguida.

A partir da avaliação acima, sugere-se como medidas para a RMQAP: reduzir a frequência de coleta das estações JC e PC e, se for necessário mais cortes, manter apenas a estação Poço Casinho (PC) e realocar Poço Tarzan de forma a torná-la representativa da bacia agrícola, função que não está desempenhando. No caso de recursos para serem aplicados na rede sugere-se incluir outra estação em uma localidade diferente para tentar melhorar o ajuste do modelo para a área próxima a JC.

Com relação à bacia urbana, as estações não apresentaram um ajuste satisfatório ao modelo de um modo geral, tanto para vazão quanto para o parâmetro de qualidade nitrato. Vale observar que os usuários da bacia não foram incluídos nessa primeira aplicação da metodologia e que maioria deles fica localizada nessa área. Talvez As estações Morin (M) e Liceu (L) são as que oferecem maior impacto na calibração do modelo de vazões.. A calibração de Liceu (L) oferece um bom ajuste aos dados de Esperança (E), de acordo com o cenário III. Por outro lado, isso não ocorre em relação a Morin (M) que tem seu ajuste impactado nesse cenário. Morin (M) fica localizada num afluente do rio principal, enquanto Esperança (E) e Liceu (L) ficam no rio Piabanha (Tabela 2). É verdade que Liceu (L) consegue capturar parcialmente o comportamento de M, tendo em vista que no cenário II o ajuste de M é bem inferior. Assim como, os ajustes de Liceu (L) e Esperança (E).

Para a calibração da qualidade de água observa-se um impacto maior da estação Liceu. Morin, ao contrario da situação anterior, tem erros relativos baixos e com pouca variação dentre os cenários. Os valores do I_{RMQAP} demonstram o maior impacto da estação Liceu e o menor impacto da estação Esperança, mesmo esta última tendo área incremental maior do que a de Morin. Assim, avaliando a RMQAP na bacia urbana chegaríamos a seguintes conclusões: a estação Esperança (E) poderia ter sua frequência reduzida ou ser descontinuada. No caso de recursos para investimento poderia ser instalada mais estações na bacia para tentar melhorar o ajuste do modelo.

Por último, é possível notar um comportamento particular da estação Rocio em relação às demais. Vale ressaltar que sendo a estação da bacia preservada ela exerce uma função de estação *background* ou de base para avaliação dos impactos de uso do solo. Ela oferece impacto significativo na calibração de vazões e um bom ajuste de ambos os modelos. Ela pode ser parcialmente representada pela estação Pedro do Rio (PR), a partir dos resultados do cenário II para vazões. Por isso, seu I_{RMQAP} a colocou

como a terceira estação mais importante da bacia só ficando atrás das duas estações com as maiores áreas. Assim ela precisa ser mantida na rede.

Vale observar, ainda, uma prevalência do impacto da calibração das vazões no cálculo I_{RMQAP} a partir dos valores de F1 e F2 na Tabela 33 com exceção de Liceu devido ao ótimo ajuste das redes neurais aos dados de vazão. Isso significa que a partir dos dados de vazão podem ser gerados, com certa acurácia, os dados de nitrato. Dessa forma, a calibração da qualidade de água não precisa ser priorizada para avaliação da RMQAP. Dessa forma, as redes neurais mostraram ser uma ferramenta eficiente, também, para a avaliação da RMQAP. Podendo, assim, ser usadas de forma conjunta com o modelo hidrológico, principalmente, nos locais onde o ajuste do modelo de qualidade de água não é bom, mas o de vazão apresenta bom desempenho. Como, por exemplo, no caso das estações Parque Petrópolis e Pedro Rio.

6 Conclusões e Recomendações

Nesta tese, foram propostas duas metodologias para a avaliação da RMQAP a fim de selecionar estações e/ou parâmetros relevantes para o monitoramento. A primeira metodologia tem como base a análise de componentes principais não-linear desenvolvido a partir de redes neurais autoassociativas. A segunda metodologia empregou redes neurais para a formulação de um índice de avaliação de impacto dos dados observados da estação de monitoramento da RMQAP na calibração do modelo SWAT.

Nesta tese, foram propostas duas metodologias para a avaliação da RMQAP a fim de selecionar estações e/ou parâmetros relevantes para o monitoramento. A primeira metodologia tem como base a análise de componentes principais não-linear desenvolvido a partir de redes neurais auto associativas. A segunda metodologia empregou redes neurais para a formulação de um índice de avaliação de impacto dos dados observados da estação de monitoramento da RMQAP na calibração do modelo SWAT.

A realização de uma contextualização do tema do monitoramento de qualidade de água no Brasil e no mundo foi muito importante para entender a relevância do desenvolvimento dessas ferramentas.

Para o desenvolvimento das metodologias foi realizado um mapeamento das principais técnicas de avaliação de RMQA. Observou-se que o uso de uma variedade de técnicas em diferentes aplicações torna extremamente complicado definir um procedimento geral de avaliação de RMQA que atenda às particularidades de cada região. Nesse mapeamento, identificou-se a ampla utilização da análise de componentes principais linear e algumas aplicações de modelagem hidrológica com os modelos SPARROW e SWAT para avaliação de RMQA. Além disso, foram observadas significativas aplicações de Redes Neurais em dados de qualidade de água bem como recomendações para o seu uso na área de avaliação de redes.

Também foi possível fazer um levantamento do histórico e da situação atual da pesquisa em bacias experimentais e representativas e demonstrar a importância dos avanços no tema de qualidade de água. Faz-se necessário o fortalecimento da estrutura

de suporte institucional das bacias experimentais brasileiras por meio dos órgãos públicos de fomento, redes de universidades e outras instituições.

A área de estudo desse trabalho são as bacias experimentais e representativas inseridas na bacia do Rio Piabanha, cuja RMQA é operada pelo Serviço Geológico do Brasil – CPRM. Assim, foi feita uma caracterização da área de estudo necessária para o desenvolvimento das ferramentas de avaliação da RMQAP. Nessa caracterização observou-se que o objetivo da RMQAP era a avaliação dos impactos dos usos do solo relativos às bacias experimentais sendo que a operação da rede sofre constantes restrições orçamentárias. Além disso, foram reunidas informações relativas aos dados de precipitação, qualidade de água, vazão, climatologia, bem como selecionada a cartografia necessária para as aplicações das metodologias desenvolvidas nessa tese.

A primeira metodologia proposta, denominada ACPNL/RNAA/IG, tem como objetivo a identificação dos parâmetros de qualidade da água e as estações de monitoramento mais relevantes da RMQAP. Esta identificação é baseada em sua redundância e é conduzida com o objetivo de selecionar parâmetros e locais de monitoramento que poderiam ser eliminados sem perda significativa de informação, ou seja, sem impacto na variabilidade dos dados.

Para este fim, a ACPNL/RNAA/IG foi definida considerando primeiro, os parâmetros de qualidade de água e, posteriormente, os locais de monitoramento como variáveis. Foi possível avaliar a redundância da rede de qualidade da água e classificar os parâmetros e os locais de monitoramento com base em sua relevância. Os resultados foram comparados aos obtidos com ACP convencional, que é a prática padrão para este tipo de problemas.

Em ambos os casos, os resultados da ACPNL/RNAA/IG e da ACP foram diferentes, e verificou-se que a ACPNL/RNAA/IG reconstruiu melhor os dados com base nos índices CTI, MSE, EPF e R² e explicou uma fração maior da variância (FEV). Esta diferença pode ser explicada pelo comportamento não-linear típico dos dados de qualidade da água que não são capturados pela ACP convencional. Se o comportamento das variáveis fosse puramente linear, os resultados deveriam ter sido equivalentes. Já a ACPNL é capaz de capturar tanto comportamentos lineares quanto não lineares.

Os resultados da ACPNL/RNAA/IG mostraram que o parâmetro de qualidade da água mais relevante na bacia hidrográfica do rio Piabanha é Coliformes fecais (CF) e o

menos importante é a Demanda Química de Oxigênio (DQO). No entanto, não foi possível separar um conjunto de parâmetros com alto IG e outro com IG baixo. Já para as estações de monitoramento, as duas estações mais relevantes, com base em seus valores de IG, foram Rocio (R) e Liceu (L), e os dois menos relevantes, Pedro do Rio (PR) e Esperança (E). Os resultados encontrados sugerem uma predominância da bacia urbana na variabilidade da qualidade de água da área de estudo e a realocação de uma estação de monitoramento, Esperança.

Portanto, é possível, por exemplo, em caso de restrições orçamentárias, utilizar a metodologia para remover parâmetros a serem medidos ou realocar estações com base nos parâmetros e locais menos relevantes, respectivamente, sem perder informações importantes relativas à variabilidade da qualidade de água. Vale ressaltar que a abordagem ACPNL/RNAA/IG demonstrou a importância de monitorar cada bacia experimental para melhor capturar a variância total dos dados da rede, ou seja, mostrou que a RMQAP está parcialmente atingindo seus objetivos através das estações mais relevantes.

No entanto, recomenda-se desenvolver um software amigável para superar as limitações de compreensão da metodologia e, assim, atingir o maior número de pessoas que possam se beneficiar de tal metodologia.

O desenvolvimento da segunda metodologia teve como base a mensuração do impacto do uso dos dados das estações da RMQAP na calibração do modelo hidrológico SWAT para a simulação de vazões e do parâmetro de qualidade nitrato (NO₃). Isso porque, em locais onde o modelo oferece uma boa representação da bacia e os dados da estação não geram melhorias nas métricas estatísticas de avaliação do modelo, o monitoramento pode ser feito com uma frequência menor ou ser descontinuado no caso de restrições orçamentárias, por exemplo. Por outro lado, os dados das estações que melhoram consideravelmente o desempenho do modelo devem ser priorizados.

Para a primeira aplicação da metodologia, foram definidos cinco cenários de calibração. Os cenários foram calibrados para os dados de vazão e de do parâmetro de qualidade de água nitrato a partir do modelo SWAT. Esses cenários são relativos a agrupamentos possíveis de estações que se pretende avaliar o impacto. O impacto de cada estação foi calculado através do índice de impacto, IRMQAP, desenvolvido para a bacia que possui duas parcelas: uma referente à calibração das vazões e outra referente à

calibração do parâmetro de qualidade. Isso porque é sabido que há uma relação entre vazão e qualidade de água e, por isso, a calibração da qualidade de água no modelo SWAT, que é um modelo físico, requer que seja feita primeiramente a calibração de vazões. Baseado nisso, definiu-se que cada parcela de impacto do índice seria ponderada de acordo com o ajuste (representados pela média do R² e do NSE) dos dados de qualidade de água através de uma rede neural simples cuja entrada seria a vazão. Assim, quanto melhor o ajuste da rede neural maior a o peso do impacto da calibração das vazões no impacto total mensurado pelo índice. O IRMQAP é então calculado a partir do coeficiente NSE do cenário de maximização dividido pelo somatório dos erros relativos entre cenários para cada parcela.

A calibração do modelo SWAT teve bons resultados para vazões, mas, infelizmente, o ajuste não foi satisfatório para a maioria das estações. Por outro lado, a previsão do parâmetro nitrato para cada estação através dos seus dados de vazão, usando redes neurais, apresentou ajustes minimamente satisfatórios para toda a rede, com exceção da estação de monitoramento Liceu. Dessa forma, a RN é uma ferramenta eficiente, também, para a avaliação da RMQAP e pode ser usada de forma conjunta com o modelo hidrológico, principalmente, nos locais onde o ajuste do modelo de qualidade de água não é bom, mas o de vazão apresenta bom desempenho.

Com base nos resultados da aplicação da metodologia procedeu-se a avaliação da RMQAP, finalidade principal desse estudo, cujo o objetivo é avaliação do impacto de cada bacia experimental na qualidade de água com o mínimo de custo possível. De acordo com o IRMQAP as estações mais impactantes da bacia são Pedro do Rio (PR) e Parque Petrópolis (PP), principalmente, pelas suas grandes áreas de contribuição relativamente às demais estações. Como o índice é ponderado pela área o valor para essas estações se eleva. Entretanto, foi observado a partir dos resultados das simulações de vazão, que a estação PR representa bem a estação PP embora essa tenha sido considerada relevante pelo índice. Sendo assim, como a sua calibração de qualidade de água não apresentou bom desempenho recomenda-se, então, a redução da frequência de coletas para redução de custos e usar redes neurais de forma combinada ao modelo hidrológico para prever os dados de nitrato, por exemplo.

A bacia agrícola é a menos impactante na calibração do modelo, tendo duas das suas estações com os menores valores de IRMQAP da rede: Poço Tarzan e João Cristh (JC). Já Poço do Casinho (PC) possui um impacto relevante. Assim, sugere-se como

medidas para a RMQAP: reduzir a frequência de coleta das estações JC e PC e, se for necessário reduzir custos, manter apenas a estação Poço Casinho (PC) e realocar Poço Tarzan de forma a torná-la representativa da bacia agrícola, função que não está desempenhando bem. No caso de recursos para serem aplicados na rede sugere-se incluir outra estação em uma localidade diferente para tentar melhorar o ajuste do modelo para a área próxima a João Cristh (JC).

As bacias urbana e preservada são as mais impactantes. Com relação à bacia urbana, as estações não apresentaram um ajuste satisfatório ao modelo tanto para vazão quanto para o parâmetro de qualidade nitrato, embora duas delas, Morin (M) e Liceu (L), possuam valores de IRMQAP consideráveis. Vale observar que os usuários da bacia não foram incluídos nessa primeira aplicação da metodologia e que maioria deles fica localizada nessa área. Recomenda-se a utilização dos dados desses usuários para aplicações futuras. Assim, avaliando a RMQAP na bacia urbana concluímos que a estação Esperança (E) poderia ter sua frequência reduzida ou ser descontinuada. No caso de recursos para investimento poderia ser instalada mais estações na bacia para tentar melhorar o ajuste do modelo.

Quanto a Rocio, a única estação da bacia preservada, foi possível perceber que ela está capturando bem o comportamento da sua bacia experimental. Além disso, obteve um bom ajuste em relação às duas calibrações e um impacto significativo ficando em terceiro lugar no ranking de IRMQAP.

Para melhorar o desempenho do modelo SWAT na bacia sugere-se uma pesquisa mais abrangente acerca dos parâmetros de calibração que podem estar relacionados a esta bacia. Talvez um maior detalhamento do mapa de solos possa contribuir para a eficiência do modelo.

A aplicação dessa metodologia como ferramenta de avaliação da RMQAP proporcionou uma ampla investigação da relação das estações entre si, entre as estações e a bacia experimental e entre elas e o modelo, principalmente, pelo SWAT ser um modelo físico.

Finalmente, é importante mencionar que estavam disponíveis apenas três (para a primeira metodologia) e quatro anos (para a segunda metodologia) de dados de qualidade de água e que um período de registro mais longo pode agregar mais informações a respeito das correlações entre as variáveis. Além disso, um conjunto de

dados maior permitiria a inclusão de mais parâmetros no estudo (que foram removidos na etapa de pré-processamento de dados) aumentando, assim, o escopo do estudo.

Seria interessante, também, o desenvolvimento de um programa que automatizasse essa avaliação combinando as diferentes etapas e facilitando a sua aplicação pelos gestores de rede.

Recomenda-se a aplicação da metodologia para as escalas diária e sub-diária assim que houver disponibilidade desses dados na bacia dada as pequenas dimensões da bacia representativa, e, assim, comparar com os resultados da escala mensal.

Para esta primeira avaliação não foram realizados estudos relacionados a quantificação de incertezas. Recomenda-se a sua realização nas próximas aplicações da metodologia 2.

Sugere-se, ainda, que sejam realizadas mais aplicações de ambas as metodologias com um maior conjunto de dados para garantir maior robustez do método. Nessa primeira aplicação, recomenda-se que a avaliação da RMQAP realizada seja encarada apenas como uma orientação até que seja possível a utilização de uma serie mais longa de dados de qualidade de água para validação das metodologias propostas.

7 Referências Bibliográficas

- ABBASPOUR, K. C., VEJDANI, M., HAGHIGHAT, S. 2007. "SWATCUP calibration and uncertainty programs for SWAT". In: *Proc. Intl. Congress on Modelling and Simulation (MODSIM'07)*", 1603-1609. L. Oxley and D. Kulasiri, eds. Melbourne, Australia: Modelling and Simulation Society of Australia and New Zealand.
- ABDI, H., VALENTIN, D., EDELMAN, B., 1999. *Neural Networks*, Series: Quantitative Applications in the Social Sciences Sage Publications Papers.
- ALDRICH, C., AURET, L., 2013. "Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods". Springer Verlag London, 2013.
- Agência Nacional de Águas (ANA), 2012, *Panorama da qualidade das águas superficiais do Brasil*. Brasília: DF. Disponível em: http://pnqa.ana.gov.br/rede/rede_avaliacao.aspx. Acesso em 11-02-13.
- Agência Nacional de Águas (ANA), 2014, *Programa Nacional de Avaliação da Qualidade das Águas*. 2.ed Brasília: DF. Disponível em: <http://arquivos.ana.gov.br/institucional/sge/CEDOC/Catalogo/2013/programaNacionalAvaliacaoQualidade.pdf>. Acesso em 05-10-15.
- ALAM M.J., AND GOODALL J.L., 2012. "Toward disentangling the effect of hydrologic and nitrogen source changes from 1992 to 2001 on incremental nitrogen yield in the contiguous United States". *Water Resour. Res.*, 48, pp. 1-16.
- ALMEIDA, L., SERRA, J.C.V., 2017. "Modelos hidrológicos, tipos e aplicações mais utilizadas". *Rev. FAE, Curitiba*, 20, n. 1, pp. 129 – 137.
- ALVERGA, L., H., M., levantamento e cadastro de bacias experimentais no Brasil. 70 p. Trabalho de Conclusão de Curso. João Pessoa: Universidade Federal da Paraíba - Departamento de Engenharia Civil e Ambiental, 2016.
- ANDRADE, C.D., 2016, *Avaliação de escala de monitoramento e do comportamento hidrológico na bacia do rio Piabanha/RJ*, 184 p., Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ANDREWS, R., DIEDERICH, J., TICKLE, A. B., 1995. "Survey and critique of techniques for extracting rules from trained artificial neural networks". *Knowledge-Based Systems*, 8, pp. 373-389.
- ARAÚJO, L. M. N.; MORAIS, A.; VILLAS-BOAS, M.D. *et al.* "Estudos Integrados de Bacias Experimentais Parametrização Hidrológica na Gestão de Recursos Hídricos da Bacia do Rio Piabanha". In: *XVII Simpósio Brasileiro de Recursos Hídricos*, 11, São Paulo, 2007,. Anais... CD-ROM
- ARAÚJO, L.M.N., 2016, *Identificação de padrões hidrológicos de precipitação e de umidade do solo na bacia hidrográfica do rio Piabanha/RJ*, 250 p., Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ARNOLD, J. G., MORIASI, D. N., GASSMAN, P. W., *et al.*, 2012, "Swat: model use, calibration, and validation", *American Society of Agricultural and Biological Engineers*, v. 55(4), pp. 1491-1508.

- ARNOLD, J. G., R. SRINIVASAN, R. S. MUTTIAH, AND J. R. WILLIAMS, 1998. "Large-area hydrologic modeling and assessment: Part I. Model development." *J. American Water Resour. Assoc.*, 34, pp. 73-89.
- BERRY, Michael J. A., LINOFF, Gordon S. Data mining techniques. USA: Wiley Publishing Inc, 2004. 2ª edição.
- BEVERIDGE, D., ST-HILAIRE, A., OUARDA, T.B.M.J., 2012, "A geostatistical approach to optimize water quality monitoring networks in large lakes: Application to Lake Winnipeg", *Journal of Great Lakes Research*, v. 38, pp. 174–182.
- BOCHNER, J. K., 2010, *Proposta Metodológica para Identificação de Áreas Prioritárias para Recomposição Florestal – Estudo de Caso: Bacia Hidrográfica do rio Macacu/RJ*. Dissertação de M.Sc., UFRRJ - Instituto de florestas - Programa de Pós-graduação em ciências ambientais e florestais, Seropédica, RJ, Brasil.
- BORDEN, C. and ROY, D., 2015. "Water Quality Monitoring System Design". Publicado: International Institute for Sustainable Development, Canadá. Disponível em: <https://www.iisd.org/sites/default/files/publications/water-quality-monitoring-system-design.pdf> Acesso em: 08/08/2016.
- BRASIL. Lei nº. 9.433: Política Nacional de Recursos Hídricos. Brasília: 1997 Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L9433.htm. Acesso em: 03/06/2016
- BRESSIANI, D.A, 2016, *Coping with hydrological risks through flooding risk index, complex watershed modeling, different calibration techniques, and ensemble streamflow forecasting*, 200 p., Tese de D.Sc., Programa de Pós Graduação e área de Concentração em Hidráulica e Saneamento – Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, SP, Brasil.
- BARROS, V.R., 2012, *Avaliação da evapotranspiração utilizando o algoritmo SEBAL e imagens Landsat – TM – Estudo de Caso Bacia do Rio Piabanha/RJ*, 159 p., Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- BARRETO, L.V. *et al.* "Relationship between stream flow and water quality in a river section. Ambiente e Água" - *An Interdisciplinary Journal of Applied Science*, 1, pp. 118-129.
- BEHMEL, S., DAMOUR, M., LUDWIG, R., RODRIGUEZ, M., 2016. "Water quality monitoring strategies — A review and future perspectives". *Sci Total Environ.*, 571, pp. 1313-1329.
- BICKNELL, B.R., IMHOFF, J.C., KITTLE, J.L., DONIGIAN, A.S., JOHANSON, R.C. 1993. "Hydrological Simulation Program – FORTRAN (HSPF): Users Manual for Release 10. EPA-600/R-93/174", U.S. EPA, Athens, GA, 3060. <https://nepis.epa.gov/Exe> Acesso em: 30/04/2016.
- BROWN, L.C. AND BARNWELL, T.O., 1987. "The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS" (EPA/600/3-87-007). US Environmental Protection Agency, Athens.
- CAMBIEN, N., 2017. *Evaluation of the Soil and Water Assessment Tool (SWAT) to simulate pesticide dynamics in the Guayas River basin (Ecuador)*, 67p, Dissertação de M.Sc., Ghent University, Gent, Bélgica.

CARVALHO FILHO, A., LUMBRERAS, J.F., SANTOS, R.D., Os Solos do Estado do Rio de Janeiro - *Estudo Geoambiental do Estado do Rio de Janeiro* – Brasília: CPRM, 2000. <http://www.cprm.gov.br> Acesso em: 18/10/2013.

CATTELL, R. B., 1966. “The scree test for the number of factors”, *Multiv Behav Res*, v. 1, pp. 245-276.

CE, 2000 - DIRECTIVE 2000/60/EC OF THE EUROPEAN PARLIAMENT - Water Framework Directive. Disponível em: http://eur-lex.europa.eu/resource.html?uri=cellar:5c835afb-2ec6-4577-bdf8-756d3d694eeb.0004.02/DOC_1&format=PDF Acesso em: 10/09/16.

CE, 2016 - <https://www.eea.europa.eu/themes/water/status-and-monitoring>

CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS. *Prospecção Tecnológica em Recursos Hídricos. Qualidade da Água Superficial - Síntese de Painel de Especialistas. Anexo III-a*. Brasília: CGEE, 2003.

CETINKAYA, P., HARMANCIOGLU, N.B., 2012, “Assessment of Water Quality Sampling Sites by a Dynamic Programming Approach”, *J. Hydrol. Eng*, v.17, pp. 305–317.

CHAPMAN, D., "A Guide to Use of Biota, Sediments and Water in Environmental Monitoring - Second Edition ". Strategies for water quality assessment, 2 ed., chapter 2, In: UNESCO/WHO/UNEP, 2 ed. 1996.

CHAPMAN, D. S., PURSE, B. V., 2011 “Community versus single-species distribution models for British plants”. *Journal of Biogeography*, 38, pp. 1524–1535.

CHEN, Q., WU, W., BLANCKAERT, K., MAA, J., HUANG, G., 2012, “Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses”. *J. Environ. Manage.* 110, pp. 116-124.

CHILUNDO, M., KELDERMAN, P., ÓKEFFE, J.H., 2008, “Design of a water quality monitoring network for the Limpopo River Basin in Mozambique”, *Physics and Chemistry of the Earth*, v. 33, pp. 655–665.

CHO, H., OLIVERA, F., 2009. “Effect of the Spatial Variability of Land Use, Soil Type, and Precipitation on Streamflows in Small Watersheds”. *JAWRA*, 45, pp.673-686.

CHOI, S. W., LEE, I., 2004. “Nonlinear dynamic process monitoring based on dynamic kernel PCA”. *Chem. Eng. Sci*, v. 59, pp. 5897 – 5908.

COELHO, M., FERNANDES, C. V. S., 2015. Métodos para a Avaliação da Representatividade Espacial em Monitoramento de Qualidade da Água para a Gestão de Recursos Hídricos (Parte I). In: XXI Simpósio Brasileiro de Recursos Hídricos, 2015, Brasília. Segurança Hídrica e Desenvolvimento Sustentável: desafios do conhecimento e da gestão. Porto Alegre: ABRH, 2015. v. 1. p. 1-8. 34.

COELHO, M., FERNANDES, C. V. S., 2015. Métodos para a Avaliação da Representatividade Espacial em Monitoramento de Qualidade da Água para a Gestão de Recursos Hídricos (Parte II). In: XXI Simpósio Brasileiro de Recursos Hídricos, 2015, Brasília. Segurança Hídrica e Desenvolvimento Sustentável: desafios do conhecimento e da gestão. Porto Alegre: ABRH, 2015. v. 1. p. 1-7.

COLLISCHONN, W. ; BUARQUE, D. C.; PAIVA, R. C. D. ; FAN, F. M. ; KAYSER, R. H. B., Manual introdutório para aplicação do modelo MGB-IPH - versão 2.0. 2010.

<https://www.ufrgs.br/hge/mgb-iph/downloads/mgb-iph-2011-versao-antiga/> Acesso em 12/09/2017.

CPRM – Serviço Geológico do Brasil, Empresa Brasileira de Agropecuária – EMBRAPA, Departamento de Recursos Minerais - DRM/RJ, 1999, “*Projeto Rio de Janeiro*” acessado em Agosto de 2013.

DE CICCIO, V., 2009, Estimativa da Evapotranspiração em Mata Atlântica pelos Métodos dos Balanços Hídrico e de Cloreto. CUNHA/SP. Disponível em: <<http://www.sbpnet.org.br/livro/64ra/resumos/resumos/3780.htm>>.

DE PAULA, T.P., 2011, *Diagnóstico e Modelagem Matemática da Qualidade da Água em Trecho do Rio Piabanha/RJ*, 159 p. Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

DAVIES-COLLEY, R.J., SMITH, . G., WARD, R. C., BRYERS, G. G., MCBRIDE, G. B., QUINN, J. M.; SCARSBROOK, M. R., 2011. “Twenty Years of New Zealand’s National Rivers Water Quality Network: Benefits of Careful Design and Consistent Operation”. *JAWRA*, vol. 47, pp. 750-771.

DEVIA, K., GAYATHRI & GANASRI, B.P., DWARAKISH, G.S., 2015. “A Review on Hydrological Models”. *Aquat Procedia*. 4, pp. 1001-1007.

DO, H.T., LO, S., CHIUEH, P. *et al.*, 2012, “Design of sampling locations for mountainous river monitoring”, *Environmental Modelling & Software*, v. 27-28, pp. 62-70.

DUNTEMAN, G. H., 1989. *Principal Components Analysis*, second edition, SAGE.

EMBRAPA, 2003. *Levantamento de Reconhecimento de Baixa Intensidade dos Solos do Estado do Rio de Janeiro*. Boletim de Pesquisa e Desenvolvimento v. 32. Disponível em: <https://www.infoteca.cnptia.embrapa.br/handle/doc/1037806> Acesso em: 10/05/2014

DOGAN, E., SENGORU, B., KLUB, R., 2009. “Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique”. *J. Environ. Manage.*, 90, pp. 1229-1235.

EMBRAPA, 2013 - <https://www.cnpm.embrapa.br/projetos/relevobr/download/rj/sf-23-z-b.htm> Acesso em: 08/07/2013.

EPA, 2016 - <https://www.epa.gov/laws-regulations/summary-clean-water-act>

EVSUKOFF, A.G., 2012. Notas de Aula da Disciplina Data Science – COC800, Programa de Engenharia Civil - COPPE/UFRJ, 2012.

FAN, X., CUI, B., ZHAO, H., ZHANG, Z., ZHANG, H., 2010, “Assessment of river water quality in Pearl River Delta using multivariate statistical techniques”, *Procedia Environ. Sci.*, v. 2, pp. 1220–1234.

FERNANDEZ, G. Data Mining using SAS applications. USA: Chapman & Hall, 2003.

FUNDAÇÃO COPPETEC – COPPE/UFRJ, 2010, Relatório Técnico Parcial 2: Estudos Integrados de Bacias Experimentais – Parametrização Hidrológica na Gestão de Recursos Hídricos das Bacias da Região Serrana do Rio de Janeiro - Chamada Pública MCT/FINEP/CT-HIDRO - Bacias Representativas 04/2005.

- GIROLAMO, A.M., LO PORTO, A., PASSARELLA, G., GARNIER, M., 2003. "Evaluation of the Optimal Location of Monitoring Sites Based on Hydrologic Models and GIS Technology". TWRI Technical Report 266. 2^a Conferência Internacional do SWAT
- GITAU, M.W., CHAUBEY, I., 2010. Regionalization of swat model parameters for use in ungauged watersheds, *Water*, 2, pp. 849-871.
- GONÇALVES, R.C., 2008, *Modelagem Hidrológica do tipo chuva-vazão via SMAP e TOPMODEL - Estudo de Caso: Bacia do Rio Piabanha/RJ*. 173p., Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- GU, W., LIU1, J., LU1, J., FRENTRESS, J., 2013. "Current Challenges in Experimental Watershed Hydrology". Current Perspectives in Contaminant Hydrology and Water Resources Sustainability, Dr. Paul Bradley (Ed.), InTech, DOI: 10.5772/55087. Available from: <https://www.intechopen.com/books/current-perspectives-in-contaminant-hydrology-and-water-resources-sustainability/current-challenges-in-experimental-watershed-hydrology>. Acesso em: 10/12/20017
- GUEDES, H.; SILVA, D. D.; ELESBON, A. A. A.; RIBEIRO, C. B. M.; MATOS, A. T.; SOARES, J. H. P., 2012. "Aplicação da análise estatística multivariada no estudo da qualidade da água do Rio Pomba, MG." Revista Brasileira de Engenharia Agrícola e Ambiental - Agriambi, Campina Grande, v. 16, n. 5, pp. 558-563.
- HAN, J., KAMBER, M., PEI, J., 2011, *Data Mining: Concepts and Techniques*, 3 ed., Morgan Kaufmann.
- HARMANCIOGLU, N.B, ALPASLAN, N., 1992, "Water quality monitoring network design a problem of multi-objetive decision making", *JAWRA*, v. 28, pp. 179-192.
- HARMEL, R.D., COOPER, R.J., SLADE, R.M., et al., 2006, "Cumulative uncertainty in measured streamflow and water quality data for small watersheds". *Trans. ASABE*, v. 49 (3), pp. 689 – 701.
- HARMEL, R.D., SMITH, P.K., 2007, "Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling", *J. Hydrol.* v. 337, pp. 326–336.
- HEWLETT, J.D., LULL, H.W., and REINHART, K.G., 1969, In defense of experimental watersheds: *Water Res. Research*, v. 5, no. 1, pp. 306-316
- HASENACK, H.; WEBER, E.; MARCUZZO, S. (org.). Diagnóstico Ambiental de Porto Alegre: Geologia, Solos, Drenagem, Vegetação e Ocupação. Porto Alegre: Secretaria Municipal do Meio Ambiente, 2008. pp. 84.
- HASTIE, T., AND W. STEUTZLE, 1989. "Principal curves". *J. Amer. Stat. Assoc.*, 84, pp. 502–516.
- HAND, D.J., 1998. "Data mining: Statistics and more?" *Amer. Statistician*, 52, 112–118.
- HARMEL, R.D., COOPER, R.J., SLADE, R.M., HANEY, R.L., ARNOLD., J.G. 2006. "Cumulative uncertainty in measured streamflow and water quality data for small watersheds". *Transactions of the ASABE*, 49, pp. 689-701.
- HOROWITZ, A. J., 2013. "A review of selected inorganic surface water quality-monitoring practices: are we really measuring what we think, and if so, are we doing it right?". *Environ. Sci. Technol.*, v. 7, pp. 2471–2486.

- HOWES, P., CROOK, N., 1999, "Using input parameter influences to support the decisions of feedforward neural networks". *Neurocomputing*, v. 24, Issues 1–3, pp. 191–206pp.
- HSIEH, W. W., 2007. "Nonlinear principal component analysis of noisy data", *Neural Networks*, Vol. 20, pp. 434-443.
- INEA, 2017 – Mapa de Uso e Cobertura do Solo da Bacia do Paraíba do Sul – RJ elaborado pelo Instituto Estadual do Ambiente (INEA RJ) / Diretoria de Biodiversidade, Área Protegidas e Ecossistemas (DIBAPE) /Coordenadoria de Gestão do Território e Informações Geoespaciais (COGET) - <http://www.visualizador.inde.gov.br/> Acesso em: 01/10/2017
- IM, S. BRANNAN, K., MOSTAGHIMI, S., CHO, J., 2003. "A Comparison of SWAT and HSPF Models for Simulating Hydrologic and Water Quality Responses from an Urbanizing Watershed". ASAE Paper No. 032175. <http://swat.tamu.edu/media/90096/asae032175.pdf> Acesso em: 18/04/2016.
- JACQUES, P. D., SHINZATO, E., 2000. Uso e Cobertura do solo – Estudo Geoambiental do Estado do Rio de Janeiro Brasília: CPRM, 2000. <http://www.cprm.gov.br> Acesso em: 18/10/2013.
- JOLLIFFE, I.T., 2002. *Principal Component Analysis*, Second Edition, Springer.
- JÖRESKOG, K. G., 1967. "Some Contributions to Maximum Likelihood Factor Analysis." *Psychometrika*. Vol. 32, 4, pp. 443–482.
- KAISER, H. F., 1960. "The Application of Electronic Computers to Factor Analysis". *Educ. Psychol. Meas.*, v. 20, pp. 141-151.
- KANTARDZIC, M., 2011, *Data Mining: Concepts, Models, Methods, and Algorithms*, 2 ed., John Wiley & Sons.
- KHALIL, B., OUARDA T.B.M.J., ST-HILAIRE, A., Chebana, F., 2010. A statistical approach for the rationalization of water quality indicators in surface water quality monitoring networks. *J.Hydrol.* v. 386, pp. 173–185.
- KHALIL, B., OUARDA, T.B.M.J., 2009, "Statistical approaches used to assess and redesign surface water-quality-monitoring networks", *J. Environ. Monitor.*, v. 11, pp. 1915-1929.
- KOHONEN, T., 1995. "Self-Organizing Maps", *Springer Series in Information Sciences*, Vol. 30, Springer, Berlin, Heidelberg, New York, 1995.
- KRAMER, M. A., 1991, "Nonlinear Principal Component Analysis using Autoassociative Neural Networks", *AIChE J.*, v. 37, pp. 233-243.
- KRAMER, M. A., 1992. "Autoassociative Neural Networks". *Comput. Chem. Eng.*, v. 16, pp. 313-328.
- KRUGER U., ZHANG J., XIE L., 2008. "Developments and Applications of Nonlinear Principal Component Analysis – a Review." In: Gorban A.N., Kégl B., Wunsch D.C., Zinovyev A.Y. (eds) *Principal Manifolds for Data Visualization and Dimension Reduction. Lecture Notes in Computational Science and Enginnee*, vol 58. Springer, Berlin, Heidelberg
- KUNWAR, P. S., MALIKA, A., MOHANA, D., S., SARITA, 2004. "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) - a case study", *Water Res.*, v. 38, pp. 3980–3992.

- LETTENMAIER, D.P., ANDERSON, D.E., Brenner, R.N., 1984. "Consolidation of a stream quality monitoring network", *Water Resour. Bull.*, v. 20, pp.473-481.
- LEOPOLD, L. B., 1970, "Hydrologic research on instrumented watersheds" *Internat. Assoc. Sci. Hydrol.*, no.97, pp. 135-150.
- LEE, C.J, HIRSCH, R. M., SCHWARZ, G.E., HOLTSCHLAG, D.J., PRESTON, S.D., CRAWFORD, C.G., VECCHIA, A.V., 2016. "An evaluation of methods for estimating decadal stream loads". *J. Hydrol.*, 542, 185-203.
- LO, S., KUO, J.T., WANG, S., 1996. "Water Quality Monitoring Network Design of Keelung River, Northern Taiwan". *Wat Sci Technol*, 34. 49-57.
- LOU, R.F., 2010, *Modelagem Hidrológica Chuva-vazão e Hidrodinâmica Aplicada na Bacia Experimental do Rio Piabanha/RJ*, 174p., Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- LOUCKS D.P., VAN BEEK, E., 2017. *Water Quality Modeling and Prediction*. In: Water Resource Systems Planning and Management. Springer, Cham.
- LU, B.W., PANDOLFO, L., 2011, "Quasi-objective nonlinear principal component analysis", *Neural Networks*, v. 24, pp.159–170.
- MAASSEN, S., BALLA, D., KALETTKA, T. *et al.*, 2012, "Screening of prevailing processes that drive surface water quality of running waters in a cultivated wetland region of Germany - A multivariate approach", *Sci Total Environ*, v. 438, pp. 154–165.
- MACKAY, D. J. C., 1992, "Bayesian Interpolation". *Neural Comput* ,v. 3, pp. 415–447.
- MAGYAR, N., HATVANI, I. G., SZÉKELY, I. K., *et al.*, 2013, "Application of multivariate statistical methods in determining spatial changes in water quality in the Austrian part of Neusiedler See, *Ecol.Eng.*, v. 55, pp. 82–92.
- MAHJOURi, N., KERACHIAN, R., 2011, "Revising river water quality monitoring networks using discrete entropy theory: the Jajrood River experience", *Environ. Monit. Assess.*, v. 175 , pp. 291-302
- MAIER, H. R., JAIN, A., DANDY, G.C., *et al.*, 2010, "Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions". *Environ. Modell.Softw.*, v. 25, pp. 891-909.
- MAINARDI, F., 2013. *Simulação dos impactos de lançamentos de poluentes sobre a qualidade da água de bacias hidrográficas integrada com sistema de informação geográfica*, 207p., Dissertação de M.Sc., Programa de Pós Graduação em Engenharia de Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.
- MASCARENHAS, F. C. B. Projeto – EIBEX-I Estudos Integrados de Bacias Experimentais – Parametrização Hidrológica na Gestão de Recursos Hídricos das Bacias da Região Serrana do Rio de Janeiro. Rio de Janeiro: MCT/FINEP/CT-HIDRO, 2007. 72 p.
- MCBRIDE, G. B., 2005, *Using statistical methods for water quality management: issues, problems and solutions*, 1 ed., Wiley Interscience.
- MIRGOLBABAEI, H., ECHEKKIA, T., SMAOUIB, N., 2014. "A nonlinear principal component analysis approach for turbulent combustion composition space". *Int. J .Hydrogen Energ*, v. 39, pp. 4622-4633.

- MISHRA, A., 2010, "Assessment of Water Quality Using Principal Component Analysis: A Case Study of the River Ganges", *J. Water. Chem. Technol.*, v. 32, pp. 227–234.
- MONAHAN, A. H., 2001. "Nonlinear principal component analysis: Tropical Indo-Pacific sea surface temperature and sea level pressure", *J. Clim.*, v.14, pp. 219–233.
- MONAHAN, A. H., 2000. "Nonlinear principal component analysis by Neural Networks: Theory and Applications to Lorenz System", *J. Clim.*, v.13, pp. 821–835.
- MONTENEGRO, A.A.A., DE LIMA, J.L.M.P., ABRANTES, J.R.C.B.O., SANTOS, T.E.M., 2013. Impact of mulching on soil and water conservation in semiarid catchment: Simulated rainfall in the field and in the laboratory. *Bodenkultur*, 64, pp. 79–85.
- MONTENEGRO, A.A.A., MONTENEGRO, S.M.G.L., LIMA, J.E.F.W., KOIDE, S., ARAGÃO, R., SOUZA, C., CRUZ, M.A.S., 2014. Hydrological studies in experimental and representative basins in Brazil: the experience of the REHIDRO network. 15th Biennial Conference ERB2014.
- MORASI, D. N., ARNOLD, J., G., VAN LIEW, M.W., *et al.*, 2007, "Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations". *Trans. ASAE*, v. 50(3), pp. 885–900.
- MEDEIROS, V. M. A.; MOREIRA, L. F. F.; RIGHETTO, A. M. "Modelagem do processo chuva-vazão na bacia experimental de Serra Negra do Norte. RN" In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 16., 2005, João Pessoa. Anais...João Pessoa: ABRH, 2005.
- NACE, R. L., 1969, *Water and man: a world view; the International Hydrological Decade*. UNESCO and its programme. Paris, UNESCO, 46 p.
- NASA, 2017 - <https://www2.jpl.nasa.gov/srtm/> Acesso em: 08/11/2017.
- NAJAH, A., EL-SHAFIE, A., KARIM, O. A. EL-SHAFIE, A.H., 2013. "Application of artificial neural networks for water quality prediction". *Neural. Comput. Appl.*, pp. 187–201.
- NCEP/CFSR, 2013 – Texas A&M Agrilife Research. Global Weather Data for SWAT. College Station, TX - <https://globalweather.tamu.edu/> - Acesso em 20/09/2013.
- RNH, 2013 – Banco de dados Hidroweb - <http://hidroweb.ana.gov.br/default.asp> Acesso em: 16/09/2013.
- INMET, 2013 – <http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesautomaticas> Acesso em 9/10/2013
- NEITSCH, S. L., ARNOLD, J. G., KINIRY, J. R., AND WILLIAMS, J. R., 2011. "Soil and Water Assessment Tool Theoretical Documentation Version 2009", Grassland, Soil and Water Research Laboratory, Agricultural Research Service and Blackland Research Center, Texas Agricultural Experiment Station, College Station, Texas.
- NGUYEN, D., WIDROW, B., 1990, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights" *Proceedings of the International Joint Conference on Neural Networks*, v. 3, pp. 21–26.

NIRAULA, R., NORMAN, L.M., MEIXNER, T., CALLEGARY, J.B. 2012, “Multi-gauge Calibration for modeling the Semi-Arid Santa Cruz watershed in Arizona-Mexico border area using SWAT” *Air, Soil and Water Research*, 5, pp. 41-57.

NRCS, 2007. Part 630 Hydrology National Engineering Handbook - Chapter 7

Hydrologic Soil Groups. Disponível em:

<https://directives.sc.gov.usda.gov/OpenNonWebContent.aspx?content=17757.wba>.

Acesso em: 12/12/2017.

OLSEN, R. L., CHAPPELL, R. W., LOFTIS, J. C., 2012, “Water quality sample collection, data treatment and results presentation for principal components analysis – literature review and Illinois River watershed case study”. *Water Res.*, v. 46, pp. 3110–3122.

OUYANG, Y., 2005, “Evaluation of river water quality monitoring stations by principal component analysis”, *Water Res.*, v. 39, pp. 2621-2635.

OUYANG, Y., NKEDI-KIZZA, P., WU, Q.T. *et al.*, 2006, “Assessment of seasonal variations in surface water quality”, *Water Res.*, v. 40, pp. 3800 – 3810.

PAGE, R. M., LISCHIED, G., EPTING, J., *et al.*, 2012, “Principal Component analysis of time series for identifying indicator variables for riverine groundwater extraction management”, *J.Hydrol.*, v. 432–433, pp. 137–144.

PAIVA, J.B.D.de, PAIVA, E.M.C.D., 2001. *Hidrologia aplicada á gestão de pequenas bacias hidrográficas*. Porto Alegre: ABRH., 625 pp.

PAPADOKONSTANTAKIS, S., LYGEROS, A., JACOBSSON, S.P., 2006, “Comparison of recent methods for inference of variable influence in neural networks”. *Neural Networks*, v. 19, pp. 500–513.

PARAJULI, P.B., OUYANG, Y, 2013. “Watershed-Scale Hydrological Modeling Methods and Applications, Current Perspectives in Contaminant Hydrology and Water Resources Sustainability” Dr. Paul Bradley (Ed.), InTech.

PARK, S.-Y., J.H., WANG, S., *et al.*, 2006, “Design of a water quality monitoring network in a large river system using the genetic algorithm”, *Ecol Modell.*, v. 199, pp. 289-297.

PASSARELLA, G., LO PORTO, Antonio & De Girolamo, Anna & Garnier, Monica. (2003). Evaluation of the optimal location of monitoring sites based on hydrologic models and GIS technology.

PIMENTEL DA SILVA, L.; ROSA, E.U.; SILVA, C.P.P., 2010. “Caracterização de parâmetros físicos e do saneamento ambiental de bacia experimental representativa localizada na Baixada de Jacarepaguá, Rio de Janeiro, Brasil”. *Ambi-Água*, Taubaté, v. 5, n. 3, pp. 232-244.

PORTO, M, 2003. CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS. *Prospecção Tecnológica em Recursos Hídricos. Qualidade da Água Superficial - Documento Final. Anexo II-a*. Brasília.

PURI, D., BOREL, K., VANCE, C., KARTHIKEYAN, R., 2017. “Optimization of a Water Quality Monitoring Network Using a Spatially Referenced Water Quality Model and a Genetic Algorithm”. *Water*, pp. 704-715.

PRIESLEY, C.H.B., TAYLOR, R.J., 1972. On the assessment of surface heat flux and evaporation using large scale parameters, *Monthly Weather Rev.*, 100, n.2, pp. 81-92.

- QIU, L.J., ZHENG, F.L., YIN, R. S., 2012. "SWAT-based runoff and sediment simulation in a small watershed, the loessial hilly-gullied region of China: capabilities and challenges". *Int. J. Sediment Res.*, 27, pp. 226-234.
- RACHWAL, M.F.G., CURCIO, G.R., DEDECEK, R.A., 2008, "A influência das características pedológicas na produtividade de acácia-negra (*Acacia mearnsii*), Butiá, RS Pesquisa Florestal Brasileira", *Colombo*, v.56, pp.53-62.
- RAZAVI, T. COULIBAL, P., 2013, "Classification of Ontario watersheds based on physical attributes and streamflow series." *J.Hydrol.*, v. 493, pp. 81-94.
- RAZMKHAH, H., ABRISHAMCHI, A., TORKIAN, A., 2010, "Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran)", *J. Environ. Manage.*, v. 91, pp. 852-860.
- RESINO, S., SEOANE, J.A., BELLON, J.M, DORADO, J., SANCHES-MARTIN, F., ALVAREZ, E., COSIN, J., LOPEZ, J. C., LOPEZ, G., MIRALLES, P., BERENGUER, J. (2011). "An artificial neural network improves the non-invasive diagnosis of significant fibrosis in HIV/HCV coinfecting patients". *Journal of Infection*, 62, pp. 77-86.
- RIBEIRO, C. B. M.; BRESSIANI, D. A. ; ROTUNNO FILHO, O. C. . Modelagem hidrológica de vazões na bacia Amazônica utilizando o modelo SWAT. In: XXI Simpósio Brasileiro de Recursos Hídricos, 2015, Brasília. Segurança Hídrica e Desenvolvimento Sustentável: desafios do conhecimento e da gestão, 2015.
- RODRIGUES, L. N., "Instrumentação e monitoramento de uma bacia hidrográfica experimental da Rede AgroHidro". SEMINÁRIO DA REDE AGROHIDRO, 1., .. Rio de Janeiro. Água: desafios para a sustentabilidade da agricultura: anais. Rio de Janeiro: Embrapa Solos, 2014. 83 p. (Embrapa Solos. Documentos, 167) pp. 49-51.
- ROSÁRIO, L. S., 2013, *Análise das áreas de proteção permanente com suporte de geotecnologias face a possíveis alterações no código florestal brasileiro – estudo de caso: bacia do rio Piabanha/RJ*. XVI, 146, Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- ROSS, P. J. VALDES AND S. WIGGINS, 2008. "ENSO dynamics in current climate models: an investigation using nonlinear dimensionality reduction". *Nonlin. Processes Geophys.*, 15, pp.339–363.
- RUNKEL, R. L., C. G. CRAWFORD, AND T. A. COHN. 2004. Load estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers. U.S. Geological Survey Techniques and Methods Book 4.
- RUMELHART, D.E., HINTON, G.E., WILLIAMS, R.J., 1986, "Learning representations by back-propagating errors" *Nature*, v. 323, pp. 533-536.
- RUSSELL, S.J., NORVIG, P. (Stuart Jonathan) "Artificial intelligence: a modern approach"/ Stuart Prentice Hall, Englewood Cliffs, New Jersey, 1995.
- SANTHI, C., KANNAN, N., ARNOLD, J. G., DI LUZIO, M., 2008. "Spatial calibration and temporal validation of flow for regional-scale hydrologic modeling". *J. American Water Resour. Assoc.* 44, pp. 829-846.
- SARTORI, A., LOMBARDI NETTO, F. GENOVEZ, A.M., 2005, "Classificação Hidrológica de Solos Brasileiros para a Estimativa da Chuva Excedente com o Método

do Serviço de Conservação do Solo dos Estados Unidos Parte 1: Classificação” *RBRH – Revista Brasileira de Recursos Hídricos*, v. 10 (4), pp. 05-18.

SARKAR, A., PANDEY, P., 2015, “River Water Quality Modelling Using Artificial Neural Network Technique”. *Aquatic Procedia*, 4, pp.1070-1077.

BOWERS, J.A., SHEDROW, C.B., 2000. Predicting stream water quality using artificial neural networks. U.S. Department of Energy Report WSRC-MS-2000-00112, 7 pp.

SAXTON, K.E., RAWLS, W.J. , 2006. ” Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions” *Soil Science Society of Agronomy Journal*, 70, pp. 1569-1578.

SWAT, 2012. SWAT2012 Input/Output File Documentation – Manual Disponível em: <http://swat.tamu.edu/documentation/> Acesso em: 15/09/2013.

SCANNAPIECO, D., NADDEO, V., ZARRA, T., *et al.*, 2012, “River water quality assessment: A comparison of binary- and fuzzy logic-based approaches”, *Ecol Eng.*, v. 47, pp. 132–140.

SCHOLZ, M., 2012 “Validation of nonlinear PCA”, *Neural. Process. Lett.*, v. 36, pp. 21-30.

SEO, I.W, YUN, S. H., HUN, S.Y., 2016. “Forecasting Water Quality Parameters by ANN Model Using Pre-processing Technique at the Downstream of Cheongpyeong Dam”. *Procedia Engineering*, 154, pp. 1110-1115.

SELLE, B., SCHWIENTEK, M., LISCHIED, G., 2013, “Understanding processes governing water quality in catchments using principal component scores”, *J.Hydrol*, v. 486, pp. 31-38.

SHARMA, S., ISIK, S., SRIVASTAVA, P., 2013. “Deriving spatially distributed precipitation data using the artificial neural network and multilinear regression models”, *J. Hydrol. Eng.*, v.18, pp.194-205.

SHARP, W.E., 1971. “A Topologically Optimum Water-Sampling Plan for Rivers and Streams”, *Water Resour. Res.*, v.7, pp. 1641-1646.

SCHWARZ, G., HOOS, A., B ALEXANDER, B., SMITH, R., 2006. “The SPARROW Surface Water-Quality Model: Theory, Application and User Documentation”. U.S. Geological Survey Techniques and Methods, Livro 6, B, Capítulo 3. https://pubs.usgs.gov/tm/2006/tm6b3/PDF/tm6b3_titlepages.pdf Acesso em: 01/07/2016.

SCHMITT, J., 2005, *Pré-processamento para a mineração de dados: uso da análise de componentes principais com escalonamento ótimo*, 144 p., Dissertação de M.Sc., Universidade Federal de Santa Catarina, SC, Brasil.

SHRESTHA, S., KAZAMA, F., 2007, “Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan”, *Environ.Modell.Softw.*, v. 22, pp. 464-475.

SILVA, A. C., 2014. *Simulação da descarga fluvial em resposta a mudanças de uso e cobertura da terra: bacia do rio Bonfim, Petrópolis (RJ)*, 124 p., Dissertação de M.Sc., Universidade Federal do Rio de Janeiro – UFRJ, Programa de Pós-Graduação em Geografia, RJ, Brasil.

- SILVA, F.H.B.B., 2000. Método de determinação de Escoamento em bacias Hidrográficas a partir de levantamentos pedológicos. Rio de Janeiro: Embrapa Solos v. 21.
- SILVA, J.G.P., VILLAS-BOAS, M.D., KENUP, R. E., GOULART, D. R., MEDEIROS, A. D., SANTOS, F.J., HENRIQUES, M. P., MELO, J.S., RIBEIRO, M. J., Avaliação preliminar das séries temporais de precipitação dos 8 anos de monitoramento na bacia experimental e representativa do Rio Piabanha – região serrana do Rio de Janeiro-RJ. In: *XXII Simpósio Brasileiro de Recursos Hídricos*, Florianópolis, 2017.
- SIMEONOVA, V., STRATISB, J.A., SAMARAC, C., *et al.*, 2003. “Assessment of the surface water quality in Northern Greece”, *Water Res.*, v. 37, pp. 4119–4124.
- SMITH, R., SCHWARZ, G., ALEXANDER, R., 1997. “Regional Interpretation of Water-Quality Data”. *Water Resour. Res.*, 33, 2781–2798.
- SINGHAL, A., SALSBUURY, T.I., 2005. “A simple method for detecting valve stiction in oscillating control loops”, *J. Process Control*, v. 15, pp. 371-382.
- SOARES, P. F., 2001. *Projeto e Avaliação de Desempenho de Redes de Monitoramento de Qualidade da Água utilizando o Conceito de Entropia.*, 170p., Tese de D.Sc., USP, São Paulo, SP, Brasil.
- ŠRAJ, M.; RUSJAN, S.; VIDMAR, A.; MIKOŠ, M.; GLOBEVNIK, L.; BRILLY, M., 2008. “The experimental watersheds in Slovenia”. *Earth Environ. Sci.*, v. 4 . 1315_4_1_012051.pdf>. Acesso: 12 dez. 2010.
- STROBL, R.O., ROBILLARD, P.D., 2008, “Network design for water quality monitoring of surface freshwaters: A review”, *J. Environ. Manage.*, v. 87, pp. 639–648.
- SUN, S., FU, G., DJORDJEVIC, S., KHU, ST., 2012. “Separating aleatory and epistemic uncertainties: probabilistic sewer flooding evaluation using probability box”, *Journal of Hydrology*, v. 420-421, pp. 360-372.
- TETRA TECH, Inc. 2004, Progress Report Development of Nutrient Criteria in California: 2003-2004. Prepared for US EPA Region IX. Tetra Tech, Lafayette, CA.
- TESHAGER, A.D., GASSMAN, P.W., SECCHI, S., SCHOOOF, J.T., MISGNA, G., 2016. “Modeling agricultural watersheds with the Soil and Water Assessment Tool (SWAT): Calibration and validation with a novel procedure for spatially explicit HRUs”. *Environ. Manage.*, 57, pp. 894-911.
- TELICI, I.T., NAM, K., GUAN, J., *et al.*, 2009. “Optimal water quality monitoring network design for river systems”, *J. Environ. Manage.*, v. 90, pp.2987-2998.
- THEODORIS, S., KOUTROUMBAS, K., *Pattern Recognition*, 4a edition, Academic Press, 2009.
- TOEBES, C., OURYVAEV, V., 1970. *Representative and Experimental Basins, A International Guide for Research and Practice*. Studies and Reports in Hydrology, 4. UNESCO, Paris. 348 pp.
- TUCCI, C.E.M. Modelos hidrológicos. Porto Alegre, Ed. Universidade UFRGS, 1998.
- USGS, 2016 - <https://earthexplorer.usgs.gov/> Acesso em: 14/08/2016.

- UUSITALO, L., LEHIKONEN, A., HELLE, I. et al., 2015. “An overview of methods to evaluate uncertainty of deterministic models in decision support”, *Environmental Modelling & Software*, v. 63, pp. 24-31.
- VANDENBERGHE, V., VAN GRIENSVEN, A., BAUWENS, W., 2001. “Sensitivity analysis and calibration of the parameters of ESWAT: Application to the river Dender”, *Water Science & Technology*, v.43 (7), pp. 295-301.
- VANDENBERGHE, V., VAN GRIENSVEN, A., BAUWENS, W., 2002. “Detection of the most optimal measuring points for water quality variables: application to the river water quality model of the river Dender in ESWAT”. *Water Science & Technology*, v. 46 (3), pp. 1-7.
- VANDENBERGHE, V., VAN GRIENSVEN, A., BAUWENS, W., 2007. “Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns.” *Environmental Modelling & Software*, v. 22, pp.725-732.
- VANDENBERGHE, V., VAN GRIENSVEN, A., BAUWENS, W., et al., 2003. “Propagation of uncertainty in diffuse pollution into water quality predictions: application to the river Dender in Flanders, Belgium”. In: Proceedings of the Seventh International Specialised Conference on Diffuse Pollution and Basin Management, 17e22 August 2003, Dublin, Ireland.
- VEGA, M., PARDO R., BARRADO E. et al., 1998, “Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis”, *Water Res.*, v. 32, pp. 3581–3592.
- VICENS, G.J., RODRIGUEZ-ITURBE, I., SHAAKE, J.C., 1975, “A Bayesian framework for the use of regional information in hydrology” *Water Resour. Res.*, v. 11 (3), pp. 405–414.
- VILLAS-BOAS, M.D., BASTOS, A.O., ARAÚJO, L.M.N., F.J., SILVA, J.G.P., MONTEIRO, A.E.G.C, 2011, “Manejo do uso do solo como mecanismo regulatório da gestão da qualidade da água - estudo de caso: a bacia do rio Piabanha”. In: XIVth IWRA World Water Congress, Setembro de 2011, Porto de Galinhas, IWRA.
- VILLAS-BOAS, M.D., OLIVERA, F., AZEVEDO, J.P.S., “A water quality index model using stepwise regression and neural networks models for the Piabanha River basin in Rio de Janeiro, Brazil”, Abstract H23A-1229, *2013 Fall Meeting, AGU*, San Francisco, Calif., 9-13 December 2013.
- VILLAS-BOAS, M.D., F.J., SILVA, J.G.P., HENRIQUES, M. P., RIBEIRO, M. J., KENUP, R. E., MEDEIROS, A. D., AZEVEDO, J.P.S, OLIVERA, F., Os 10 anos do projeto institucional da CPRM: estudos integrados em bacias experimentais e representativas – região serrana/RJ– conquistas e desafios futuros. In: *XXII Simpósio Brasileiro de Recursos Hídricos*, Florianópolis, 2017.
- Villas-Boas, M.D., Olivera, F., AZEVEDO, J. P., 2017. “Assessment of the water quality monitoring network of the Piabanha River experimental watersheds in Rio de Janeiro, Brazil, using autoassociative neural networks”. *Environ. Monitor Assess.* 189, pp. 439-454.
- WARD, R. C., 1996. “Water Quality Monitoring: Where’s the Beef?”, *JAWRA*, vol 32, pp. 673-680.
- WARD, R.C., LOFTIS, J.C., MCBRIDE, G.B., In: John Wiley & Sons, *Design of Water Quality Monitoring Systems*, Hoboken, New Jersey, 2003.

- WASSERSTEIN, R., 2010. "George Box: a model statistician". *Significance*, 7, pp. 134-135.
- TUCCI, C.E.M., 2010. *Conceitos dos Modelos Hidrológicos*. In. <http://rhama.com.br/blog/index.php/sem-categoria/conceitos-dos-modelos-hidrologicos/> Acesso em: 10/08/17
- WHITE, M.J., HARMEL, R.D., ARNOLD, J.G., WILLIAMS, J.R., 2014. "SWATCheck: a screening tool to assist users in the identification of potential model application problems". *J. Environ. Qual*, 43, pp. 208-214.
- World Health Organization (WHO), United Nations Environment Programme (UNEP), United Nations Educational, Scientific and Cultural Organization (UNESCO) and World Meteorological Organization (WMO), 1992: Global Environment Monitoring System (GEMS)/Water Operational Guide. Inland Waters Directorate, Burlington, Ontario, 3ed. Disponível em: <https://www.ircwash.org/sites/default/files/242-92GE-10022.pdf> Acesso em: 15/10/2016.
- World Meteorological Organization, 1994. Guide to hydrological practices: Data acquisition and processing, analysis, forecasting and other applications. WMO-No. 168. Geneva, Switzerland: WMO
- World Meteorological Organization – WMO, 2008. *Guide to hydrological practices: Data acquisition and processing, analysis, forecasting and other applications*. WMO-No. 168. Suíça: WMO.
- YAN AN, Y., ZOU, Z., LI, R., 2016. "Descriptive Characteristics of Surface Water Quality in Hong Kong by a Self-Organising Map", *Int. J. Environ. Res. Public Health*, 13, 115, pp. 1-13.
- YOON, Y., GUIMARAES, T., SWALES, G., 1994. "Integrating artificial neural networks with rule-based expert systems". *Decis. Support. Syst.*, 11, pp. 497-507.
- ZHANG, X., SRINIVASAN, R., VAN LIEW, M. 2008. "Multi-Site Calibration of the SWAT Model for Hydrologic Modeling". *Transactions of the ASABE*, 51., pp. 2039-2049.
- ZECKOSKI, R.W., SMOLEN, M.D. MORIASI, D.N. FRANKENBERGER, J.R., FEYEREISEN, G.W., 2015. "Hydrologic and water quality terminology as applied to modeling". *Transactions of the ASABE*, 58, pp. 1619-1635.
- ZHU, X.; YUE, Y.; WONG, P.W.H.; ZHANG, Y.; TAN, J., 2018. "Optimum Water Quality Monitoring Network Design for Bidirectional River Systems." *Int. J. Environ. Res. Public Health*, 15, pp.195-214.
- ZOBEL, C. W., COOK, D. F., 2011, "Evaluation of neural network variable influence measures for process control", *Eng. Appl. Artif. Intell.*, v. 24, pp. 803–812.